



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

December 2, 2008

DSSD 2010 CENSUS COVERAGE MEASUREMENT MEMORANDUM SERIES #2010-B-07

MEMORANDUM FOR Donna L. Kostanich, Co-Chair
Census Coverage Measurement Operational Integration Team

Magdalena Ramos, Co-Chair
Census Coverage Measurement Operational Integration Team

From: Harland H. Shoemaker, Jr. (*signed*)
Chair, Census Coverage Measurement Survey Design Subgroup

Prepared by: Michael B. Moldoff
Decennial Statistical Studies Division

Subject: The Design of the Coverage Measurement Program for the
2010 Census

The attached document on The Design of the Coverage Measurement Program for the 2010 Census is being provided as background material for the Census Coverage Measurement Estimation Workshop to be held in January 2009.

Since some of the details of the operations are subject to revision, updates to this document will be issued as necessary.

Attachment

cc:
DSSD CCM Contacts List

U S C E N S U S B U R E A U

The Design of the Coverage Measurement Program for the 2010 Census

	Page	
1.0	Introduction	1
1.1	Overview of the 2010 CCM Design	2
1.2	High-Level Schedule of 2010 CCM Activities	5
1.3	CCM Activities and Census Activities	6
2.0	CCM Sample Overview	6
3.0	Identifying CCM Sample Areas	8
3.1	Block Clustering	8
3.2	Block Cluster Selection Process	9
3.3	Small Block Cluster Subsampling	10
4.0	Independent Listing	11
5.0	Initial Housing Unit Matching and Followup	13
5.1	Initial Housing Unit Computer Matching	13
5.2	Initial Housing Unit Before Followup Clerical Matching	14
5.3	Initial Housing Unit Followup	15
5.4	Initial Housing Unit After Followup Clerical Matching	16
6.0	Identifying the Sample Housing Units	17
6.1	Identifying the P Sample	17
6.2	Identifying the PI Sample	18
6.3	Identifying the E Sample	19
7.0	Person-based Operations	19
7.1	Person Interview	20
7.2	Automated Geocoding	22
7.3	Automated Residence Status Coding	22
7.4	Person Matching Design	23
7.5	Person Computer Matching	25
7.6	Person Before Followup Clerical Matching	26
7.7	Clerical Geocoding and Clerical Residence Status Coding	26
7.8	Before Followup Technician and Analyst	27
7.9	Person Followup	27
7.10	Person After Followup Clerical Matching	29
8.0	Final Housing Unit Operations	30
8.1	Final Housing Unit Computer Processing	30
8.2	Final Housing Unit Before Followup Clerical Matching	31
8.3	Final Housing Unit Followup	31
8.4	Final Housing Unit After Followup Clerical Matching	32

	Page
9.0 Estimation	32
9.1 Missing Data for P-Sample Persons	33
9.1.1 Imputing Missing Demographic Characteristics for Persons	34
9.1.2 Household-Level Noninterview Adjustment for CCM Person Interview	34
9.1.3 Imputing the Probability of P-Sample Inclusion for Persons	34
9.1.4 Imputing the Probability of Match Status for Persons	35
9.2 Missing Data for E-Sample Persons	37
9.2.1 Missing Demographic Characteristics for E-Sample Persons	37
9.2.2 Imputing Missing Enumerations Status for Persons in the E Sample (Net Error)	37
9.2.3 Imputing Missing Enumeration Status for Persons in the E Sample (Component Error)	38
9.3 Person Net Coverage Error	39
9.4 Person Components of Coverage Error	39
9.4.1 Two-Stage Ratio Adjustment for E-Sample Persons	40
9.4.2 Erroneous and Correct Enumerations for Persons	40
9.4.3 Person Omissions and Omissions by Missed Housing Unit	40
9.5 Missing Data for P-Sample Housing Units	41
9.5.1 Imputing Missing Characteristics for P-Sample Housing Units	41
9.5.2 Imputing Missing Housing Unit Status for P-Sample Housing Units	41
9.5.3 Imputing Missing Matching Status for P-Sample Housing Units	41
9.6 Missing Data for E-Sample Housing Units	42
9.6.1 Missing Characteristics for E-Sample Housing Units	42
9.6.2 Imputing Missing Enumeration Status for E-Sample Housing Units (Net Error)	42
9.6.3 Imputing Missing Enumeration Status for E-Sample Housing Units (Component Error)	42
9.7 Housing Unit Net Coverage Error	43
9.8 Housing Unit Components of Error	43

10.0 References	44
-----------------	----

Attachments

1. High-Level Workflow of Key CCM Operations
2. Codes used in 2010 CCM Housing Unit Matching
3. Codes used in 2010 CCM Person Matching
4. Search Areas and CCM Person Matching Operations
5. Overview of CCM Estimation--Persons in Housing Units
6. Overview of CCM Estimation--Housing Units
7. CCM Glossary

The Design of the Coverage Measurement Program for the 2010 Census

1.0 Introduction

The Census Coverage Measurement (CCM) Program activities planned for the 2010 Census will provide estimates of net coverage error and coverage error components (including omissions and erroneous enumerations) for housing units and persons in housing units. These operations were designed to support the CCM goals for the 2010 Census. The 2010 CCM goals are to (Singh, 2005):

- begin producing measures of coverage error, including its components of omissions and erroneous enumerations
- produce these measures of coverage error not only for demographic groups and geographic areas, but also for key census operations
- continue to provide measures of net coverage error

It should be noted that group quarters facilities and persons residing in those facilities are not within the scope of the CCM Program for the 2010 Census. Similar to the 2000 Accuracy and Coverage Evaluation (A.C.E.), any area designated as a Remote Alaska enumeration area will be out-of-scope for the 2010 CCM.

The 2010 CCM estimation process will continue to use the dual system methodology for net coverage error estimation. This methodology has been used in the coverage measurement programs for the 1980, 1990, and the 2000 censuses (Fay et al 1988, Hogan 1993, U.S. Census Bureau 2004). However for 2010, logistic regression rather than the more traditional post-stratification methodology will be utilized for the first time to produce dual system estimates. In addition, a correlation bias adjustment will be applied to improve the estimates of net coverage error for certain person population subgroups.

Since estimation of coverage error components is new for the 2010 CCM, we will be using a more straightforward estimator for correct and erroneous enumerations—a basic inflation of ratio-adjusted survey weights. The ratio adjustments serve to reduce the variance of the estimates and also ensure that the estimates of correct and erroneous enumerations add to the total number of data-defined census enumerations overall and for various subpopulations. Estimates of omissions will be obtained by adding the estimate of erroneous enumerations to the dual system net error estimate.

The purpose of this document is to describe the design of the 2010 CCM Program. Subsequent sections of the document present in greater detail the features of each activity. Section 2 provides an overview of the CCM sample. Section 3 describes how the CCM sample areas are identified. Section 4 details the Independent Listing. Section 5 details the Initial Housing Unit Matching and Followup Operations. Section 6 describes how the sample housing units are identified. Section 7 details the person-based operations, while section 8 covers the Final Housing Unit Operations. Section 9 covers Estimation. An overview of the design of the 2010 CCM Program follows.

There are seven attachments. Attachment 1 is a high-level workflow of key CCM operations. Attachment 2 shows the codes used in housing unit matching, and Attachment 3 shows the codes used in person matching. Attachment 4 displays the search areas in relation to CCM Person Matching operations. Attachments 5 and 6 give an overview for CCM Estimation operations for persons in housing units, and housing units, respectively. Attachment 7 is a glossary of CCM terminology.

1.1 Overview of the 2010 CCM Design

This section presents an overview of the key 2010 CCM activities shown in Attachment 1.

The Population sample (P sample) and Enumeration sample (E sample) are the two samples used for dual system estimation (DSE) employed by the CCM program. The E sample and the dual system estimates are used to derive the components of coverage error. The P sample is derived from an independent listing of housing units (completely separate from the census) conducted in a probability sample of block clusters in each state, the District of Columbia (DC), and Puerto Rico. The primary sampling unit, as well as the basic work unit for CCM, is the block cluster. A block cluster consists of one or more geographically contiguous census collection blocks. The source of the E sample is the census housing units and census person enumerations in housing units geocoded to the sample of block clusters selected for the P sample.

The CCM design in Puerto Rico is the same as the CCM design in the U.S. stateside. All CCM data collection instruments and forms will be translated into Spanish for use by Spanish-speaking interviewers and respondents in the U.S. and in Puerto Rico. The forms developed for use in Puerto Rico will account for the special features of addresses there.

In the CCM block clusters selected in the first phase of sampling, the Independent Listing (IL) is conducted by personal visit using the CCM Independent Listing Book (ILB). This operation lists all housing units or potential housing units in each CCM block cluster. Information regarding the number of housing units belonging to a particular housing unit structure is also collected. A dependent quality control procedure is performed subsequently on a randomly selected set of addresses in each block cluster, again using the CCM ILB. Listings in block clusters that fail this quality inspection will be corrected. If it is determined that the wrong block cluster was listed, the ILB is returned to the original lister who will conduct the listing in the correct block. The ILBs are sent to the National Processing Center (NPC) for keying.

CCM IL results are used to stratify and decide the subsampling rates for small block clusters. Small block clusters are subsampled in order to minimize variance on the coverage estimates and increase field efficiency.

The Initial Housing Unit (HU) Computer Matching follows, in which the keyed results from the CCM IL are matched against a preliminary listing of census housing unit (and group quarters) addresses within each sample block cluster and one ring of surrounding blocks.¹ Duplicate

¹ A surrounding block is one that touches the block cluster at one or more points. This definition includes the blocks that touch the corner of the block cluster.

addresses within the CCM IL and within the preliminary census address listing are also identified.

The next operation is the Initial HU Before Followup (BFU) Clerical Matching, which attempts to resolve possible matches, nonmatches, and duplicates from the Initial HU Computer Matching. A computer-assisted clerical matching system is used to facilitate this operation. In addition, a clerical search is conducted for duplicate addresses within the CCM IL and preliminary census address lists. None of the results from any of the Initial HU operations are used to update any census information.

Once these initial matching activities are completed, the Initial HU Followup (IHUFU) attempts to clarify any differences between the CCM Independent Listing results and the census preliminary address file, and also to resolve potential duplicates identified in the computer or clerical matching. An attempt is also made to resolve housing unit status where information is either missing or incomplete.

When the IHUFU is completed, the Initial HU After Followup (AFU) Clerical Matching ensues, in which the results from the IHUFU are used to attempt to clarify the match status, duplicate status, or housing unit status for unresolved addresses. This process produces files that contain match codes for CCM and census housing units in the sample block clusters and in the surrounding blocks.

The next CCM operation is to identify the addresses where the CCM Person Interview (PI) will be conducted. The address list produced after the Initial HU AFU Clerical Matching provides the universe from which the PI addresses will be selected. The PI workload has two components: those addresses eligible for and selected to be in the P sample and those addresses not eligible to be in the P sample. These latter units are also referred to as “census-only” units.² For the selected housing units in each sample block cluster, interviewers employ a computer-assisted data collection instrument on a laptop to obtain information about the residents of the sample housing unit on Interview Day (the day the PI is conducted) and persons who moved out of the sample housing unit between Census Day (April 1, 2010) and Interview Day.

Including the census-only housing units and persons who moved out of the sample housing unit between Census Day and Interview Day in the PI reduces followup (since these could eventually be selected for the E sample). It also facilitates matching to the census enumerations for estimating component errors. The PI includes a quality control reinterview (PI RI) as well, which again uses an automated data-collection instrument on a laptop.

The 2010 CCM is employing the Post Enumeration Survey (PES) B+ methodology, whereby Interview Day residents of the sample housing unit and certain people who moved out of the sample housing unit between Census Day and Interview Day are included in the P sample. The 2000 A.C.E. employed the PES C methodology where the P sample included Census Day

² Units eligible to be in the P sample must have been listed during the CCM IL operation. Census-only units are units in the sample block cluster that are determined to be valid housing units missed in the CCM IL or actually located outside the sample block cluster. Census-only units are not eligible for the P sample but are likely to be in the E sample.

residents of the sample housing unit (nonmovers and outmovers), as well as inmovers. Only the Census Day residents of the housing unit were eligible for matching, however.

The identification of E-sample housing units is done by mapping the final set of census housing units onto the previously-identified CCM segments. Overlapping the P and E samples improves the precision of the dual system estimates and the cost effectiveness of succeeding field operations.

Before the Person Matching operations begin, an automated operation to assign a residence status code to all persons listed in the PI is conducted. Also, an automated operation to assign geocodes to alternate and inmover addresses³ collected in the PI is conducted. Separate clerical operations are conducted to 1) assign geocodes to respondent-provided alternate addresses and 2) assign residence status codes to persons when these codes could not be assigned during the automated coding operations.

For 2010, the CCM Person Computer Matching will attempt to search for matches between persons rostered at the sample addresses during the PI and persons enumerated in the census in the CCM sample block clusters and surrounding blocks. Alternate and inmover addresses collected during the PI are also used to identify other places in which to search for matches between PI and census people. The computer matching will also conduct a nationwide search for matches between the PI and census enumerations. A computerized search for duplicate persons in either the PI or census is also conducted. For PI, the search is within the sample block clusters. For duplicate persons enumerated in the census, we search both within the sample block clusters and either nationwide for U.S. residents or within Puerto Rico for Puerto Rico residents.

The Person BFU Clerical Matching follows, in which the matching staff utilize computer-assisted clerical matching techniques, accompanied by CCM and census maps, to review and assign the status of match, possibly match, or not match to PI and census person records utilizing the results of the previous computer matching. A clerical search for duplicates is also performed. Cases remaining unresolved in terms of match status, enumeration status, or residence status become eligible for the next CCM operation, Person Followup.

During CCM Person Followup (PFU), interviewers in the field contact cases identified in the Person Matching operations with unresolved residence or enumeration status, even when they may have been matched, in order to resolve the particular issue. Telephone interviews will also be conducted from a Census Telephone Center to followup on any possible matches or possible duplicates where one of the records in the linked pair is located outside the CCM sample block cluster or ring of surrounding blocks. The field PFU also includes a reinterview for quality control purposes. The PFU interview and reinterview both are completed using a paper questionnaire.

³ An alternate address is any address other than the CCM sample address where someone may have lived and been enumerated by the census. An inmover address is an address where someone lived before they moved into the CCM sample address.

After PFU is complete, the Person AFU Clerical Matching applies the information obtained from that interview to attempt to resolve match, residence, or enumeration status for persons sent to PFU.

Once all of the person-based operations have been performed, a series of final housing unit procedures is conducted. The first of these tasks is the Final HU Computer Processing. In this operation, a determination is made as to which housing units will proceed to the Final HU BFU Clerical Matching operation, based on changes to the lists of CCM or census units since the Initial HU Matching and IHUFU operations were conducted. No computer matching is performed during the Final HU Computer Processing.

During Final HU BFU Clerical Matching, the staff uses computer-assisted clerical matching techniques, as well as CCM and census maps, to review and assign the status of matched, possibly matched, and not matched to addresses sent from the Final HU Computer Processing. The Final HU Clerical Matching operation is similar to the Initial HU Clerical Matching operation described previously, except that they use final housing unit census data rather than data from the preliminary list of census addresses. Also, results from the Person Matching and PFU operations are available.

Those cases that need additional information are candidates for the Final HU Followup (FHUFU) operation. In this operation, attempts are made to collect information to resolve any remaining differences between the P- and E-sample housing units. The FHUFU data collection forms and the related reinterview forms are paper-based.

The concluding step in the Final Housing Unit operations is the CCM Final HU AFU Clerical Matching. When conducting this operation, results from the FHUFU (and any previous CCM Person and Housing Unit Matching operations) are used to resolve any outstanding issues regarding match, housing unit, or duplication status. Once Final HU AFU has concluded, the CCM Estimation process begins.

The CCM Estimation process consists of several operations, which will ultimately lead to the production of estimates of coverage errors for both housing units and persons in housing units. This includes estimates of net coverage error as well as estimates of the component errors, including omissions and erroneous enumerations. As part of this estimation, we will implement operations to account for missing data and reduce the sampling and nonsampling errors in the coverage estimates.

1.2 High-Level Schedule of 2010 CCM Activities

A high-level schedule of the 2010 CCM activities described in the previous section is as follows:

Table 1: High-Level Schedule of 2010 CCM Activities

Activity	Start Date	End Date
Create and Select Sample Block Clusters	October 2008	April 2009
Conduct Independent Listing	August 2009	December 2009
Conduct Small Block Cluster Subsampling	January 2010	February 2010
Conduct Initial Housing Unit Computer Matching	January 2010	March 2010
Conduct Initial Housing Unit Before Followup Clerical Matching	February 2010	March 2010
Conduct Initial Housing Unit Followup	March 2010	April 2010
Conduct Initial Housing Unit After Followup Clerical Matching	March 2010	May 2010
Identify Person Interview Sample	May 2010	June 2010
Conduct Person Interview	August 2010	October 2010
Conduct Clerical Geocoding and Residence Status Coding	October 2010	December 2010
Identify E Sample	December 2010	December 2010
Conduct Person Computer Matching	December 2010	January 2011
Conduct Person Before Followup Clerical Matching	January 2011	February 2011
Conduct Person Followup	January 2011	March 2011
Conduct Person After Followup Clerical Matching	February 2011	April 2011
Conduct Final Housing Unit Computer Processing	March 2011	April 2011
Conduct Final Housing Unit Before Followup Clerical Matching	April 2011	May 2011
Conduct Final Housing Unit Followup	May 2011	June 2011
Conduct Final Housing Unit After Followup Clerical Matching	May 2011	July 2011
Conduct CCM Estimation	May 2011	October 2012

1.3 CCM Activities and Census Activities

Attachment 1 details the relationship between the aforementioned CCM activities and 2010 Census activities. CCM uses the results of three census processes. First, in order to conduct the Initial HU Computer Matching and conduct the subsampling for small block clusters, CCM relies on the preliminary census address list. Second, to identify the E sample and conduct person matching and PFU, CCM requires the unedited person data files obtained from the census (Census Unedited File or CUF). Lastly, CCM needs census edited person and housing unit data (the Census Edited File or CEF) to create the estimates of coverage error. Shaded items in the flowchart indicate that the CCM operation occurs in the field.

In addition, to ensure independence between census and CCM, the PI operation is scheduled to start after the Census Coverage Followup operation is complete, although some minor overlap is a possibility.

2.0 CCM Sample Overview

The CCM sample design is a general-purpose sample to support the various objectives of the program, which includes the new objective of estimating erroneous enumerations and omissions in addition to net error for the 2010 Census. The sample design for 2010 is essentially the same as that used for the 2000 A.C.E., except for the double sampling reduction of medium and large

block clusters.⁴ A description of the sample design of the 2000 A.C.E. can be found in U.S. Census Bureau (2004).

The CCM sample consists of two parts. The Population Sample, P sample, and the Enumeration Sample, E sample, have traditionally defined the samples for dual system estimation. Both the P sample and the E sample measure the same housing unit and household population. However, the P-sample operations are conducted independently of the census. The E sample consists of housing units and census person enumerations in housing units in the same sample areas as the P sample. After matching with the census lists and reconciliation, the P sample provides information about the population missed in the census, whereas the E sample provides information about erroneous census inclusions. This information is used in different ways to estimate the net coverage error and the components of coverage error (see Sections 9.3 and 9.4 for more details).

The CCM sample design comprises a number of distinct processes from forming block clusters, creating the sampling frame, selecting sample block clusters, to eventually selecting addresses for the P and E samples. After the CCM block clusters are selected, an address list is created independent of the census for each CCM sample block cluster (this is the IL operation described earlier). The approximate CCM listing workload is 12,500 block clusters for the U.S. and 600 for Puerto Rico. Overall, approximately one million housing units are listed (950,000 in the nation and 50,000 in Puerto Rico). Finally, after selecting the CCM P-sample addresses for interviewing, the 2010 CCM sample is approximately 300,000 housing units for the U.S. and 15,000 for Puerto Rico (see Table 2). The national sample is distributed among the 50 states and DC roughly proportional to population size, although there are slight increases in the sample for small states and for American Indian Reservations.

Table 2 shows the expected size of the 2010 CCM sample in terms of the targeted number of housing units and sample block clusters.

Table 2: 2010 CCM Target Sample Sizes

Geography	Expected Independent Listing Sample Size (HUs)	Expected Number of Sample Block Clusters for Independent Listing	Target Number of Housing Units for the P sample
U. S.	950,000	12,500	300,000
Puerto Rico	50,000	600	15,000

⁴ The original 2000 coverage measurement survey design called for a sample about twice as large as the A.C.E., since the intent of the survey was to produce a “one-number census” which was corrected for net coverage error. A January 1999 Supreme Court ruling against the use of sampling for apportionment resulted in a reduction of the sample size. The double sampling reduction was a step necessary to accomplish this reduction and thus, not needed for 2010.

The CCM sample selection has three phases:

- Phase 1: Independent Listing (IL) Sample Selection.** The selection of the IL sample is an initial sample of approximately 13,100 block clusters for which field staff develops an independent list of housing unit addresses.
- Phase 2: Small Block Cluster Subsampling.** From the initial CCM block cluster sample, we select a subsample of small block clusters using the updated count of housing units in these block clusters.
- Phase 3: Housing Unit Identification.** This phase occurs after the Initial Housing Unit operations are completed. We identify the specific housing units in the P and E samples and the housing units that will receive the PI. The P and E samples are identified at separate times. All identification involves selecting a subsample of housing units within large block clusters. The P sample is a component of the PI sample.

3.0 Identifying CCM Sample Areas

The first two phases of the sample selection, which identify the areas in which CCM operations will be conducted, are described in this section. These two phases include three activities: block clustering, selecting the block clusters where the CCM IL operation will take place, and taking a subsample of small block clusters after the IL operation occurs.

3.1 Block Clustering

Block clustering is the first major CCM operation in 2010. The block cluster is the CCM primary sampling unit. The block clustering universe is the 50 states and DC except for remote areas of Alaska.⁵ The universe also includes the entire land area of Puerto Rico.

Block clusters are formed by combining neighboring 2010 census collection blocks, excluding those census collection blocks consisting entirely of water. Each block cluster consists of one or more geographically contiguous census collection blocks grouped together to form an average of 30 housing units. We attempt to make the land area for each block cluster reasonably compact so it can be covered by an interviewer in the field without incurring unreasonable travel costs. When formed, these block clusters may be standard city blocks or irregularly shaped areas defined for the most part by identifiable geographic boundaries. The block clustering process maintains the following general goals (Whitford, 2008):

- forming block clusters that are compact and have identifiable boundaries, to the extent possible
- forming block clusters that average 30 housing units
- minimizing the number of block clusters with 80 or more housing units, to the extent possible
- reducing the number of block clusters with no housing units, to the extent possible

⁵ For the remainder of this document, use of the word “state” includes DC.

- limiting the land area of block clusters to 15 square miles, to the extent possible
- respecting certain geographical boundaries (i.e., state, county, Local Census Office, military bases, AIR)

The source of the addresses used to generate the tally of housing units for forming block clusters is the 2010 Address Canvassing universe.⁶ The latest address updates included in this data are the Spring 2008 Delivery Sequence File⁷ refresh and the Local Update of Census Addresses (LUCA) provisional adds. For Remote Update/Enumerate areas, which do not go to the census Address Canvassing operation, the most current housing unit counts are used. These are generally counts from the 2000 Census.

3.2 Block Cluster Selection Process

Within each state and Puerto Rico, block clusters are stratified based on the cross-classification of size, tenure (owner/non-owner or rented), and American Indian Reservation (AIR) status. First, block clusters are classified by size into three mutually exclusive groups based on the expected number of housing units within the block cluster. The three groups are: small block clusters – 0 to 2 housing units, medium block clusters – 3 to 79 housing units, and large block clusters – 80 or more housing units.

The second classification categorizes medium and large block clusters based on tenure, i.e., for the entire block cluster, the proportion of persons in households for which the housing unit is either rented or owned based on Census 2000 data. To classify block clusters into strata based on tenure, the proportion non-owner population of a block cluster determines whether the block cluster is either in the non-owner stratum or the owner stratum. Tenure is defined for the housing unit. So, if one member of the household is the owner, then all members of the household are deemed owners.

For 26 states, an AIR stratum is defined. This stratum consists of block clusters with 3 or more housing units located on an American Indian Reservation or associated trustland. Table 3 summarizes the six mutually exclusive Phase 1 sampling strata.

⁶ This is the set of addresses sent to the census Address Canvassing operation.

⁷ The Delivery Sequence File (DSF) is a file of all locations where the U.S. Postal Service delivers mail. The Census Bureau obtains regular updates of this file in order to maintain its Master Address File (or MAF).

Table 3: 2010 CCM Phase 1 Sampling Strata

Sampling Stratum	Size Definition	Tenure	AIR ¹
(1) AIR	3+ housing units	N/A	yes
(2) Small	0 - 2 housing units	N/A	N/A
(3) Medium Owner	3 - 79 housing units	less than 40% is non-owner	no
(4) Medium Non-Owner	3 - 79 housing units	at least 40% is non-owner	no
(5) Large Owner	80+ housing units	less than 40% is non-owner	no
(6) Large Non-Owner	80+ housing units	at least 40% is non-owner	no

¹Only in the 26 states with sufficient American Indian population living on an American Indian Reservation in 2000. In other states, AIRs are included in the other strata.

Allocating the Sample. The Phase 1 sample is a national sample of 12,500 block clusters: 10,000 medium and large block clusters and 2,500 small block clusters. Included in the 10,000 block clusters is a separate sample of block clusters for AIR. The number of medium and large block clusters is based on an expected average of 30 P-sample housing units per block cluster given the national sample size of 300,000 housing units.

The 10,000 medium and large block clusters are allocated to the 50 states and DC. This allocation is proportional to the population in each state with three exceptions. First, Hawaii is allocated 150 block clusters to help the reliability of the Native Hawaiian and Pacific Islander estimates. Second, each state is allocated a minimum sample size of 60 block clusters. Third, 356 block clusters are allocated to states for sampling AIR block clusters. The allocation of the 2,500 small block clusters to the states is proportional to the number of small block clusters in each state.

For Puerto Rico, the Phase 1 sample is 600 block clusters: 500 medium and large block clusters and 100 small block clusters.

Probabilities of Selection. To help support the reliability of minority and renter estimates, the Phase 1 sampling operation oversamples block clusters with higher proportions of non-owner populations. This oversampling will help increase the sample size of minorities and renters. Also, the large block clusters are selected with higher probability than medium block clusters in Phase 1 to allow for the subsampling of housing units in the large block clusters in Phase 3. This brings the overall probability of selection for housing units in large block clusters more in line with the overall probability of selection for housing units in medium block clusters.

3.3 Small Block Cluster Subsampling

Following the IL that is discussed in Section 4.0, Phase 2 of sample selection is small block cluster subsampling. Utilizing results from the IL as well as preliminary census address data,

block clusters from the original small sampling stratum (0-2 housing units) are stratified into five groups based on the updated number of housing units per cluster. The updated housing unit count for each block cluster is the larger of the number of housing units from the CCM IL operation and the number of housing units from the preliminary census address list. The subsampling is designed to minimize the impact on the variance of the coverage estimates and to increase the efficiency of the CCM field operations.

The strata are formed so that small block clusters with more than the expected number of housing units will be sampled at a higher rate to keep their weights lower. This will reduce the potential contribution from these block clusters to the sampling variance of the coverage estimates.

Table 4 displays the five strata used for small block cluster subsampling, as well as their minimum sampling rates.

Table 4: 2010 CCM Phase 2 Subsampling Strata

Stratum (Cluster size)	2010 Sampling Rates
0 to 2 HUs	1-in-10
3 to 5 HUs	1-in-4
6 to 9 HUs	1-in-2.22
10+ HUs	1-in-1
0 to 9 HUs AIR/Trustland TJSA/TDSA/ANVSA*	1-in-1

* Tribal Jurisdiction Statistical Area/Tribal Designated Statistical Area/Alaska Native Village Statistical Area

As shown in Table 4, small block clusters with 10 or more housing units are not subsampled. Also, block clusters which are part of an AIR, associated trustland, TJSA, TDSA, or ANVSA (see Table 4 caption) are not subsampled. Within each of the other three strata, block clusters are sorted and a systematic sample of block clusters is selected with equal probability within each subsampling stratum.

4.0 Independent Listing

An independent listing of all housing units and potential housing units in the CCM sample block clusters is compiled before Census Day. The lister makes a complete canvass of their assigned block cluster in order to collect the following information for each housing unit and potential housing unit:

- address (information to identify the housing unit and structure, street name, city, and zip code) for units with city-style⁸ addresses and for non-city-style when available

⁸ Form of address where house number and street name are given (e.g., 123 Blue Jay Road). Non-city-style addresses pertain to all addresses other than city-style.

- P.O. Box, rural route, and box number when available (for both non-city-style and city-style addresses)
- householder name (when occupied) and physical location description for units with non-city-style addresses
- number of housing units in any multi-unit structure that is listed, including the number of units on each floor, if applicable
- description of each building that contains housing units⁹
- status of each unit listed such as: occupied or vacant and intended for occupancy, under construction, future construction, unfit for habitation, boarded up, empty trailer lot/site, structure used for storage of household goods, or structure used for another purpose
- special features of addresses in Puerto Rico will also be captured

Independent listers will use the Independent Listing Book (ILB) and a map of their assigned area to conduct the operation. All materials used by the listers are on paper.

Listers will make two attempts at the unit before contacting a proxy and use observation only when necessary. In multi-unit structures, the lister will canvass by observation and then attempt to gather and reconcile the information from a manager who may be able to provide information on all housing units more efficiently than questioning residents of each unit directly. Also, efforts will be made to acquire maps of apartment complexes, mobile home parks, and RV parks from the managers of these facilities.

Units in RV parks, campgrounds, marinas, and other similar locations will be listed only if someone with no other usual residence occupies them at the time of listing. In addition, vacant lots in mobile home parks will be listed since they could contain a unit when the CCM Person Interview is conducted. If the lister is unsure about whether or not a unit is a housing unit, they will be instructed to list it. The assignment of listers to block clusters adheres to strict rules for independence between census and CCM operations (Shoemaker, 2008a).

Listers will place a map spot and map spot number on the CCM block map to indicate the location of each structure listed. Once the field aspect of the IL concludes, these maps, along with the ILBs, are sent to the NPC for keying and post-collection processing.

There are several quality control measures taken throughout the listing process. First, a check of the CCM paper maps is completed before the maps are used in the listing operation. The listers are also observed by their crew leaders and edits of the completed ILBs are performed.

When the lister completes a block cluster, another lister, specifically hired for quality control, will check the accuracy of the original listing, as well as the map-spotted maps. For each block cluster, a random entry will be chosen from the ILBs for each block cluster. Subsequently, the selected entry and the next 11 consecutive lines in the ILB, beginning with the randomly chosen line, will be checked for accuracy using the same procedures the original listers used. If one or more of the 12 lines have been incorrectly listed, the block cluster fails the quality check and is

⁹ Description of a building includes identifying the building as: a single-family house; building with two or more apartments; mobile home or trailer; camper, tent, boat, van; or other.

reworked. All changes from the quality check are recorded. Approximately 15.8 percent of all addresses listed are expected to be subject to verification under this plan (Kostanich, 2006). The target AOQL (Average Outgoing Quality Limit)¹⁰ is 3.0 percent.

When the field-based quality checks are completed and the ILBs are sent to the NPC for keying, a quality check of the keyed data is conducted. The quality check includes two independent keying operations. The results of each are compared and differences are reconciled.

5.0 Initial Housing Unit Matching and Followup

The CCM Initial HU operations consist of several procedures, all of which are designed to assign housing unit status and match status. The Initial HU Computer Matching and Initial HU Before Followup (BFU) Clerical Matching, attempt to match units from the IL to preliminary census addresses from the Universe Control and Management System (UCM). The results of these two matching processes are recorded so that units from the IL and census units that remain unresolved can be sent to the next operation, Initial HU Followup (IHUFU). In IHUFU, information is collected that may explain any differences between the IL results and the preliminary census addresses from UCM, while also attempting to resolve potential duplicates. This information is used during the Initial HU After Followup (AFU) Clerical Matching to assign final match codes.

5.1 Initial Housing Unit Computer Matching

During Initial HU Computer Matching, all addresses listed in the CCM sample block clusters during IL are computer matched to preliminary census addresses (of housing units and GQs) within the sample block clusters and one ring of surrounding blocks. This matching operation will identify (Shoemaker, 2008b):

- matches and possible matches between the IL addresses and census addresses
- potential CCM IL duplicates
- potential census duplicates

The keyed data from the ILBs and the preliminary census addresses go through a series of data preparation steps before matching, including address standardization. Addresses from either file that are blank or could not be standardized are not computer matched, but these addresses are matched clerically.

Potential duplicates are identified through a one-to-many matching process. Links between a single census address and multiple IL addresses identify a potential CCM duplicate; links between a single CCM IL address and multiple census addresses identify a potential census duplicate.

¹⁰ The Average Outgoing Quality (AOQ) is the average quality of outgoing product including all accepted lots plus the rejected lots after all the rejected lots have been 100-percent inspected and the defectives replaced by non-defective items. The Average Outgoing Quality Limit (AOQL) is the maximum of the AOQs for all possible incoming qualities for a given sampling plan.

The results of the computer matching and additional information are loaded into the Housing Unit Matching Review and Coding System (HUMaRCS),¹¹ which are subsequently utilized by clerical matchers. A list of the match codes assigned to housing units can be found in Attachment 2.

5.2 Initial Housing Unit Before Followup Clerical Matching

The Initial HU BFU Clerical Matching procedure uses the results of the computer matching stage in an attempt to further match and review housing units from both the CCM IL and the census preliminary address file (UCM). There are two stages of Initial HU BFU Clerical Matching: BFU Technician (BFT) and BFU Analyst (BFA).

In the BFT stage, technicians will use the HUMaRCS clerical matching software to assist in the process. During this operation, the technicians will:

- review the results, including possible matches, from the Initial HU Computer Matching operation,
- attempt to match IL units and census units that were not matched in the Initial HU Computer Matching operation,
- search for duplicates:
 - between IL units within the CCM sample block clusters, and
 - between units on the UCM list in the CCM sample block cluster and the surrounding blocks, and
- review IL units for any errors (e.g., addresses that may have been keyed incorrectly, CCM map spot numbers which are out of sequence)

The supplemental materials provided to facilitate the clerical matching include images of the map-spotted CCM and census maps and other auxiliary materials, as well as location descriptions. Given the number of non-city-style addresses in rural areas, map spots and descriptions are very important for the clerical matching procedure in these areas.

In the BFA stage, analysts perform the quality control of the technicians' work and review cases referred by the technician. The referred cases are IL or census units where the match or duplicate status is not clear. The quality control process for clerical matching consists of randomly selected block clusters where a sample of the cases in the block cluster coded by the technicians will be reviewed by the analysts. For the Initial HU BFU Clerical Matching operation, approximately 16.7 percent of all block clusters going through the operation will be checked. The target AOQL is 4.0 percent.

Some block clusters meeting specific criteria will be allowed to skip the BFU Clerical Matching and go directly to IHUFU. Other block clusters will be allowed to skip both BFU Clerical

¹¹ The Housing Unit Matching and Review Coding System (HUMaRCS) is a computer-assisted clerical matching system, developed by Gunnison Consulting Group, Inc. for CCM Housing Unit Clerical Matching.

Matching and the IHUFU.¹² Some block clusters skipping BFU Clerical Matching will be subject to a clerical duplicate search in Initial HU AFU Clerical Matching.¹³

The product of the BFU Clerical Matching process will be a results file containing the initial match codes for all housing units (both IL and census), and flags indicating which census or IL units will go to the next operation, IHUFU.

5.3 Initial Housing Unit Followup

For the IHUFU operation, units from both the IL and census (UCM) that meet pre-determined criteria are followed up in the field in an attempt to determine housing unit status and/or match status. The following are the conditions that will render a unit eligible for followup (Shoemaker, 2008b):

- census unit not linked to an IL unit in the same CCM sample block cluster
- IL unit not linked to a census unit in the CCM sample block cluster or surrounding blocks
- IL or census unit with a link assigned a possible match status
- IL unit with unresolved housing unit status (under construction, future construction, unfit for habitation, vacant trailer pad, or other), regardless of match status
- IL unit with a duplicate link, regardless of match status
- IL unit with a link to a census unit in a surrounding block, regardless of match status
- census unit with a duplicate link, regardless of match status
- IL unit with a link to a census GQ

In order to maintain the independence between the CCM and census, previous census and CCM work assignments are taken into account when making assignments to CCM staff. As with all other CCM followup activities, staff will employ a docuprinted paper form seeking to acquire the necessary information. Also, paper maps with map spots from the CCM and census will be available to assist in the process.

The followup forms are customized for each individual situation. Efforts will be made to collect the following information:

- Block number (for a unit that could be a duplicate in a surrounding block).
- Evidence to determine if there was a housing unit at the address on the date of the followup visit.¹⁴ If the case in question is not a housing unit, a statement should be provided as to the reason for this determination.
- Whether or not two units identified as possible matches are the same unit.
- Whether or not two or more units identified as possible duplicates are the same unit.
- Housing unit status updates for certain situations.

¹² One example of the skip criteria: If all IL and all census addresses in a block cluster have a match code of “match” and all of the IL units have a unit status code of “occupied or vacant intended for occupancy,” the block cluster can skip both the Initial HU BFU Clerical Matching and the IHUFU.

¹³ A clerical search for duplicates in AFU will not be done for block clusters that skip BFU Clerical Matching and do not require any IHUFU field work.

¹⁴ It is possible that some of the followup cases were not housing units at the time of listing. In this situation, the interviewer will need to determine if a housing unit does actually exist at the address.

Also, interviewers will use the reference list¹⁵ provided to determine if any units sent for followup match other units on the reference list. This is done to reduce the number of potential E-sample nonmatches. Cases lacking sufficient information that would allow the field staff to locate the unit in the field and conduct the followup interview will not go to IHUFU.

The IHUFU is subject to a quality control procedure. Both data from the IHUFU interviewing process and any CCM maps that were updated are eligible for the quality control process. The process consists of field observations of the interviewers by their crew leaders, edits of completed forms and maps, and a check of data accuracy of randomly selected IHUFU forms in each CCM block cluster. A separate staff assigned to quality control operations will complete these tasks. Approximately 13.5 percent of all cases sent to IHUFU will be verified, with a target AOQL of 6.7 percent.

5.4 Initial Housing Unit After Followup Clerical Matching

Once the IHUFU is completed, the IHUFU paper forms, along with Initial HU BFU Clerical Matching results file and updated CCM maps, are made available for the next CCM operation, the Initial HU AFU Clerical Matching. There are three stages in the Initial HU AFU Clerical Matching operation: AFU Technician (AFT), AFU Analyst (AFA), and AFU Outlier Review (AFO).

A main objective of AFT is to have the technicians assign match status and/or housing unit status codes to cases that were sent to IHUFU. A clerical search is performed to attempt to find duplicates for block clusters that have skipped Initial HU BFU Clerical Matching.

In addition, address corrections and map spot number updates identified in the IHUFU operation are entered into HUMaRCS. Corrections are made to the CCM map-spotted maps as necessary and census (UCM) units not originally included on the census map-spotted maps are added and assigned a map spot number. Note that these updated census maps are used only for the CCM operations that follow; the corrections are not fed back into actual census operation.

The AFA stage for Initial HU AFU Clerical Matching is very similar to that for Initial HU BFU Clerical Matching. In this stage, the analyst conducts a quality check on a sample of technicians' work and reviews any cases referred by the technician. The expected quality control workload and target AOQL for this operation are the same as for Initial HU BFU Clerical Matching.

In the AFO stage, analysts review clusters identified for review by Headquarters staff and clusters that exceed a pre-specified threshold of weighted nonmatch and erroneous enumeration counts. The analysts review the clusters for systematic errors or issues. They document their findings in journals that can later be used during estimation processes if questions arise.

¹⁵ The reference list is a list of all of the CCM housing units and census (UCM) housing units and GQs in the CCM sample block cluster. It also includes census units in surrounding blocks which contain a unit linked to a unit in the sample block cluster.

The output for the Initial HU AFU Clerical Matching operation consists of the results file, containing the match status codes and housing unit status codes of all census and IL units included in the Initial HU operations.

6.0 Identifying the Sample Housing Units

Identifying the P- and E-sample housing units are separate activities done at different times. This is Phase 3 of the CCM sampling process. Along with identifying the P-sample housing units, an additional sample of census-only housing units is identified to receive the PI. Since these census-only units were not included in the CCM IL they cannot be part of the P sample (this would violate the independence assumption), but a sample of them is selected to be included in the PI, along with the P-sample housing units. Identifying the P-sample, the E-sample, and the census-only parts of the PI sample all involve selecting a subsample of housing units within large block clusters.

In large block clusters (80 or more housing units), the housing units within the block cluster are subsampled. This feature of the sample design is intended to increase the number of block clusters in sample while still attaining the targeted number of housing units for the P sample. The objective is to reduce costs and yield manageable field workloads for CCM PI and PFU without seriously affecting the precision of the CCM estimates. By taking advantage of the homogeneity of units within the large block clusters, interviewing a manageable fraction of several different large block clusters provides a more geographically diverse sample without significantly increasing the variance of the coverage estimates.

In Phase 1 of sampling, large block clusters have a higher selection probability than medium block clusters to take into account the anticipated subsequent housing unit subsampling. After the subsampling of housing units in large block clusters, the housing unit selection probabilities in medium and large block clusters in the same Phase 1 sampling strata are similar.

Another important goal of the housing unit subsampling is to geographically overlap the P and E samples to reduce the E-sample PFU workload. An overlapping P and E sample is not necessary but improves the precision of coverage estimates, the cost-effectiveness of the succeeding field operations, and the data processing efficiency.

6.1 Identifying the P Sample¹⁶

The CCM independently-listed units that are determined to be valid, unique housing units after the Initial HU Matching and Followup operations are the source of the P sample. For block clusters containing 79 or fewer housing units, all the CCM housing units are included in the P sample. For block clusters with 80 or more housing units, a subsample of CCM IL housing

¹⁶ Because the 2010 CCM is using the PES B+ procedure, there is a difference between the P-sample housing units and P-sample persons. The “+” in the B+ means that people who moved out of the sample address between Census Day and Interview Day into units which have no probability of selection for the P sample (GQs for instance) are included in the P sample of persons, even though their current residence can not be in the P sample of housing units. The discussions in Sections 6.0 through 6.3 focus on identifying the CCM housing unit samples only.

units is selected. All housing units in block clusters in the AIR stratum are included in the P sample.

We subsample housing units within the large block clusters by forming groups of adjacent housing units, called segments, and selecting one or more segments for the P sample. The segments have roughly equal numbers of housing units within a block cluster. Segments of housing units are used as the sampling unit in order to obtain compact interviewing workloads and to facilitate the identification of an overlapping E sample. A systematic sample of segments is selected across the block clusters within a stratum. The resulting sample will yield at least one segment selected per block cluster. The CCM IL units in block clusters where no subsampling occurs and the CCM IL units that are retained after the subsampling, comprise the P sample. The target number of block clusters and housing units to be included in the P sample are shown in Table 2.

6.2 Identifying the PI Sample

The PI sample is the set of housing units where the CCM PI is performed. It consists of the housing units selected for the P sample (described in the previous section) and a sample of census-only units. Census-only housing units are eligible for the CCM PI to obtain Census Day residence status of persons in these housing units prior to identifying E-sample housing units. There are three categories of census-only housing units eligible for the PI:

- nonmatched census units that are confirmed to exist in the sample block cluster, but missed by the CCM IL.
- nonmatched census units that are geocoded to the CCM sample block cluster, but confirmed to exist in one of the blocks surrounding the sample block cluster,
- nonmatched census units with an unresolved housing unit status code

Census-only housing units are eligible for the E sample if they are on the final census address list for the sample block cluster.

For PI sample identification, the eligible census-only housing units are assigned to the same segments described in Section 6.1 by sorting all CCM IL housing units and eligible census-only housing units together geographically by their expected location on the ground. Sometimes there are a large number of these census-only cases within a sample segment. If there is a large number (80 or more) of these census-only housing units in all the sample segments in the block cluster, they too are subsampled. Different sampling rates are used to maintain a manageable field workload. The rates vary from 1-in-2 to 1-in-4, depending on the number of census-only units in the sample segment.

If there are fewer than 80 eligible census-only housing units within all sample segments or there are fewer than 80 of these units within a block cluster that was not subsampled during the P-sample selection, all of these eligible census-only housing units are included in the CCM PI sample. Based on the 2000 A.C.E. results, we estimate about 47,000 census-only units will be included in the PI sample, in addition to the 315,000 P-sample housing units selected in the U.S. and Puerto Rico.

6.3 Identifying the E Sample

The E sample is actually identified after the CCM PI is conducted but before the Person Matching operations begin. However, the process will be discussed here in order to provide continuity in the sampling discussion.

The E sample consists of housing units enumerated in the census in the same sample areas as the P sample. The source of the final set of census housing units is the Census Unedited File (CUF). The E sample in terms of persons consists only of data-defined census persons enumerated in E-sample housing units. To be a census data-defined person, the person record must have at least two data items filled on the census form. Name is not required for the person record to be considered data-defined, but it can serve as one of the two required items.¹⁷

One primary goal of E-sample identification is to geographically overlap the P and E samples to the extent possible. Having overlapping P and E samples reduces the E-sample PFU workload. This is especially important for the CCM in the 2010 Census cycle because the PFU data collection occurs nearly a year after Census Day. Overlapping the P and E samples also improves the precision of the dual system estimates and the cost effectiveness of succeeding field operations.

The identification of E-sample housing units is done by mapping the final set of census housing units onto the previously-identified CCM segments. The final census housing units are divided into two categories; those that match to CCM IL units and those that match to census-only units. Since a sample of these census-only housing units will have gone out for a PI, we would like to maximize the probability of selecting these units for the E sample. Both types of census units found in the CCM segments selected for the P sample are eligible to be included in the E sample. If a subsample of the census-only units was drawn during PI sample identification, then only those census-only units that were selected for the PI are included in the E sample.

In addition, the census inventory of housing units may change between the Initial HU Matching and IHUFU operations and the identification of the E sample. For example, a unit may be added to the census address list through some census operation. These added units are assigned to the already-formed segments geographically.¹⁸ If there are a large number of these units in a sample segment, a systematic subsample is drawn. The subsampling rates vary from 1-in-2 to 1-in-4. The E sample is expected to contain about the same number of housing units as the P sample.

7.0 Person-based Operations

Once the Initial HU and sampling operations have been completed, a series of person-based procedures is performed. These operations aim to create an independent roster of people who are living at each housing unit selected for the CCM PI. The person-based operations include: the PI; post-PI operations including automated geocoding and automated residence status coding;

¹⁷ In order for the name to be considered as one of the filled items, it must have a total of at least three characters in the first and last name fields.

¹⁸ Variables like census collection block, street name, house number, and MAF ID are used.

Person Computer Matching; Person BFU Clerical Matching, including clerical geocoding and residence status coding; the PFU interview; and Person AFU Clerical Matching.

In the CCM PI, a series of questions are asked to identify the Interview Day and Census Day residents of the sample address, and to attempt to determine where each of these persons should or could have been counted in the census. During the Person Computer and BFU Clerical Matching operations, the information collected in the PI is compared to information collected in the census. Discrepancies and missing information are then followed up in PFU. The AFU Clerical Matching Operation uses the PFU results to resolve the cases.

7.1 Person Interview

The CCM PI will be conducted at the selected housing units to build an independent roster of people currently living at each housing unit. The census-only housing units missed in the IL, but included on the preliminary census address lists in the sample areas will be included in the PI as an attempt to resolve their Census Day residence, rather than waiting for PFU to collect the information.

Emphasis will be placed on collecting information on the Census Day and Interview Day residence of persons associated with the housing units (nonmovers and inmovers) based on Census Residence Rules. This method, by which all persons who live at the sample address at the time of the PI are included in the P sample, is known as PES Procedure B. Information on the Census Day residence of these persons is collected to facilitate matching.

When the P sample also includes persons who lived at the sample address on Census Day but who died after Census Day or whose Interview Day residence is out of scope for the coverage measurement universe (that is, has no probability of selection for the P sample), this is known as PES Procedure B+. An example of how the PI will employ the B+ methodology is collecting Census Day information on a person who was a Census Day resident of the sample address, but who moved into a college dorm between Census Day and Interview Day. The college dorm represents a residence that is out-of-scope for CCM since it is a GQ.¹⁹

In order to maintain the independence needed to preserve the validity of the DSE methodology, the PI field operation will not start until the Census Coverage Followup operation is essentially completed.

In the PI, the interviewers' primary focus is establishing the Census Day and Interview Day residence for persons residing at the sample address at the time of interview, both nonmovers and inmovers since Census Day. Data will be collected for persons (either individuals or entire households) about whom the respondent is knowledgeable and who have moved out since Census Day. Outmover data is collected to reduce the PFU workload and to obtain an earlier interview for these possible E-sample cases, thus reducing potential recall biases. This

¹⁹ Outmovers about which the PI respondent is knowledgeable are included when they move out from the sample address into a place that has no chance for selection in the P sample. These include persons currently living in group quarters or in places outside the U.S., persons who experience homelessness, and persons who die between Census Day and Interview Day.

information also helps to identify persons who may have been enumerated at more than one location in the census. Situations where all Census Day residents of a housing unit have moved out before Interview Day present a challenge for data collection and in resolving Census Day residence status.²⁰

For each household in the PI, the data collection methods will include information to establish where each person should have been counted in the census (their Census Day address) and information on where each person could have been counted, such as in-mover and alternate addresses. These in-mover and alternate addresses will be used to assist in determining the Census Day residence status of each person. They will also be used to identify possible erroneous enumerations (including possible duplicates).

The PI will be conducted using a computer-assisted person interview (CAPI) instrument on a laptop computer. Having an automated instrument allows the interview to be tailored to the specific living situation and demographic characteristics of each person, as determined by their responses. For example, the college and military questions are asked only of persons in a certain age group. A series of roster probe questions will be asked in order to capture information about any other person who has stayed at the residence that the respondent did not previously mention. These questions ask about babies, foster children, relatives, friends or someone else who may have stayed temporarily at the sample address.

After rostering, questions are asked to gather additional addresses where a person could have been counted in the census. The questions are geared towards collecting addresses at any of the following places where a person might have resided (Shoemaker, 2008c):

- Census Day address other than the sample address
- college address
- military service address
- residence for a job
- relative's address where the person may stay, including joint custody situations
- seasonal home address
- other GQ (nursing home, correctional facility, group home, etc.)
- any other place stayed often

Questions are then asked as to how often a person “cycles” (goes back and forth) between the sample address and any alternate address. This method is similar to the one employed in the 2000 A.C.E. PFU and evaluation followup interviews.

The quality control process for PI includes field observations by crew leaders, and a reinterview (PI RI) of randomly selected cases. Cases can also be identified for PI RI by either an outlier identification process or a supplemental reinterview identification process done by the quality control staff in the regional offices. An automated system called the PI RI MaRCS is used to

²⁰During residence status coding, outmovers are classified as P-sample or non-P-sample outmovers. P-sample outmovers are Census Day residents of the sample address who have moved to a different residence that cannot be counted as part of the P-sample universe. All other outmovers are non-P-sample outmovers.

select the random sample of cases and identify outliers to be reinterviewed. In the outlier identification process, data from the PI RI and the original PI are computer matched using PI RI MaRCS. When the PI and PI RI data are inconsistent, a supplemental sample of the PI interviewer's assignment is selected for PI RI.

During the PI RI operation, a quality control interviewer will visit the sample address or make contact by telephone and conduct the PI RI with the same person who responded to the PI to confirm that the interview was done. About 10 percent of the PI workload will be subject to reinterview.

7.2 Automated Geocoding

The processing of in-movers' Census Day addresses and the respondent-provided alternate addresses in the PI requires geocoding²¹ prior to automated local or nationwide matching. Since addresses are collected during both the PI and the PFU, geocoding will be required for addresses obtained by each operation B at two different times. In-mover and alternate addresses collected in the PI will be subjected to both automated and clerical geocoding, while addresses collected in PFU will only be clerically geocoded (during the Person AFU Clerical Matching operation). In-mover and alternate addresses collected in the PI in Puerto Rico will not be computer geocoded; they will only be clerically geocoded.

Automated geocoding for addresses collected in the PI will be done in time to identify census blocks for which census data will be required during the computer and clerical matching of persons to in-mover and alternate addresses (i.e., to identify the search areas). The Census Bureau's MAF/TIGER²² database is the primary source of the geocodes. Geocoding can occur at various levels of geography. The most detailed geocoding occurs when a match is made to a specific record on the MAF. If the address matches to the MAF, we actually receive the matching MAF Identifier (MAF ID) as well as the geocode. The least detailed geocode is assigning only a state. Several county codes can be assigned if only a city or town is provided and that city or town lies in more than one county. In general, geocoding is successful for matching purposes if the address can be assigned to a specific county and census block or a small group of census blocks.

7.3 Automated Residence Status Coding

Each person rostered in a PI housing unit that is coded as a complete or partial interview is assigned a residence status code. This includes people in census-only housing units selected to receive the PI. These codes ultimately are used to indicate whether a person should be in the P sample or not. The residence status codes are assigned by computer following the PI and can be reviewed and reassigned during any part of person clerical matching based on information provided in PI or PFU.

²¹ Geocoding is the process of assigning census geography to an address.

²² Master Address File/ Topologically Integrated Geographic Encoding and Referencing System. The MAF/TIGER System serves as the national repository for all of the spatial data and residential address data needed for the Census Bureau's data collection, data tabulation, data dissemination, geocoding services and map production.

The residence status codes for P-sample people are as follows (the codes are defined in Attachment 3):

- nonmover (N)
- inmover (I)
- P-sample outmover (O)
- unclassified (U)

The PI is designed to roster all persons at the sampled housing unit on Interview Day. If the respondent can provide information on people who lived at the sample housing unit on Census Day, but no longer do, then this information is collected. As such, some of the persons rostered may not turn out to be residents of the sample address on Interview Day and would therefore not be included in the P sample. The residence status codes for these non-P-sample people are as follows:

- non-P-sample outmover (Z)
- out-of-scope (X)
- never resident (A)
- unclassified outmover (C)

If the computer cannot assign a code to a particular person record, it assigns a code that flags the case for clerical review.

7.4 Person Matching Design

The proposed person matching design for 2010 differs from the 2000 A.C.E. person matching design in several important ways. Perhaps the most important difference is that the person matching and duplicate search will be expanded nationwide in 2010. This procedure will be accomplished by an automated search for matches and for duplicates across the country, as was used in evaluations of the 2000 A.C.E., (Mule 2002).

The computer-assisted clerical matching will be used to search for matches between PI person records and census enumerations within each CCM sample block cluster and surrounding blocks (this is called the “sample block cluster search area”). It will also be used to search for census enumerations of inmovers in the block cluster containing their Census Day address and in the surrounding blocks (this is called the “inmover address search area”). The inmover’s Census Day address will typically be outside the block cluster containing the sample address. A clerical search for census person enumerations where alternate address information was provided in the PI will also be conducted in the block cluster containing that address and in the surrounding blocks (called the “alternate address search area”). For possible matches and possible duplicates found in the nationwide computer search, the clerical search will be conducted only at the linked unit located outside of the sample block cluster search area (that is, the search area is limited to that specific unit). See Attachment 4 for more details on how search areas are used in the person and computer clerical matching operations.

Person matching is designed to:

- determine which PI person records in P-sample housing units belong or might belong in the P sample
- determine whether the PI person records and E-sample person records have sufficient information for followup
- determine PI person records that are duplicated within the sample block cluster search area
- determine E-sample person records that are a duplicate of a census enumeration within the sample block cluster search area or throughout the U.S.
- determine PI inmover records that link to a census enumeration within the inmover address search area
- determine PI person records that link to census enumerations within the alternate address search area
- determine PI person records found in the nationwide computer search that link to census enumerations at addresses outside the sample block cluster or surrounding blocks
- identify cases that require more information to ascertain the match status, enumeration status, and/or residence status that will be sent to PFU
- use the results of the PI and PFU interviews to assign match codes, residence status (PI persons), and enumeration status (E-sample persons)

The Person Matching operations compare people in the PI or E sample person enumerations to all census person enumerations and identify records referring to the same person. If a PI person refers to the same person as a census enumeration within the sample block cluster search area (that is, it matches), it means the person was enumerated in the census and not omitted. If an E-sample person refers to the same person as another census enumeration, it means the person was duplicated by the census.

The Person Matching operations include two primary parts: Person Computer Matching and Person Clerical Matching. The Person Computer Matching operation establishes matches when straightforward relationships between records exist. The Person BFU Clerical Matching operation allows for experienced matchers to review the computer matches and to determine match status for cases that were possibly matched or not matched by the computer.

After the Person Computer Matching and the Person BFU Clerical Matching operations are completed, the PFU operation provides more information for cases with unclear or missing data, such as match status, residence status, or enumeration status. After PFU collects more information, a Person AFU Clerical Matching operation takes place to use the supplemental information obtained via the PFU interviews.

Sections 7.5 through 7.8 and Section 7.10 provide more details on the Person Matching operations. Attachment 3 provides a list of the various codes that can be assigned during Person Computer or Person Clerical Matching.

7.5 Person Computer Matching

Computer matching for persons will be implemented using the Generalized Person Computer Matching System and related components that are used for the census unduplication processes. The Person Computer Matching in 2010 involves three different searches:

- PI to census - PI people are matched to the entire census in order to find matches both within the sample block cluster search area and throughout the census.
- E sample to census - E-sample people are matched to the entire census in order to find duplicates both within the sample block cluster search area and throughout the census.
- PI to PI within-block cluster - PI people within the block cluster are matched to find within-block cluster PI duplicates only.

For each search outlined above, Person Computer Matching for 2010 involves several steps:

1. Preparing the Data - Census and PI data fields used in the computer matching algorithm need to be consistent. These include data items such as gender and computed age. In addition, the phone number, a key component in the matching system for 2010, is edited. Finally, names are prepared by removing punctuation, titles (Mr., Mrs., Dr., etc.), and special characters (including Spanish characters). A name variant list is used to standardize first names (for instance “Cathy” and “Catherine” are considered the same).
2. Computer Matching - Person records are compared across all living quarters outside of their own (i.e., a person cannot match within their own housing unit or GQ). A person can link to multiple persons but to only one person per living quarters. Matching is accomplished in two steps. The first step matches all persons across the universe. The second step builds on the links established in the first step but provides additional flexibility for finding matches within households.
3. Geographic Assignment - Each linked pair is assigned a geographic code indicating the relationship of the link. The codes are: within the sample block cluster, in a surrounding block, in county but outside the sample block cluster and surrounding blocks, different county in the same state, or a different state.
4. Computer Modeling - Modeling is used to evaluate the links to establish which links should be considered matches, possible matches, or not a match. Rules are applied that are conditional on geographic proximity (see Geographic Assignment above), household patterns, phone numbers, person characteristics, and type of living quarters (HU or GQ). There are two types of modeling. The first is within-household modeling, which takes the household structure into consideration. The second type is residual modeling, which matches individuals without the benefit of the household structure. Within-household modeling requires a review by expert matchers to establish matching status.
5. Create the Output - The output from computer matching contains identifying information for each person record in a linked pair and the match code for that pair (either match or possible match).

Each of the three searches uses the same software and the same steps in matching. For the PI-to-census search, the data are matched only once but are modeled three times; once with respect to sample addresses, once with respect to inmover and/or alternate addresses, and once with respect to any address identified in the nationwide matching. For each of the modeling steps, the geographic distance category is reassigned based on the distance of the address being modeled. This adjustment can make it more likely to call something a match when modeling with respect to a different address. See Attachment 4 for more information on search areas used in matching.

7.6 Person Before Followup Clerical Matching

There are three phases of person clerical matching prior to PFU: Clerical Geocoding (CGC), Clerical Residence Status Coding (RSC), and Person BFU Clerical Matching (includes BFT, BFA). During CGC and RSC (which occur before Person Computer Matching), matchers (technicians and analysts) at the NPC review the results of the automated geocoding and residence status coding of the data collected in PI. During the BFU stage of Person Clerical Matching, the results of the Person Computer Matching are reviewed. All three of these phases include a technician and an analyst component.

Person Clerical Matching is assisted by the Person Matching Review and Coding software system called PerMaRCS.²³ PerMaRCS is used by NPC matchers to aid in the matching operation. PerMaRCS pre-processing is done to prepare the data for clerical matching. This pre-processing includes creating census and PI data tables, setting sufficiency and inmover address codes,²⁴ creating person links, setting person duplicate links and match codes for unlinked persons, and assigning the work units in batches before the first phase of Person Clerical Matching can begin. The PerMaRCS system also supports the Person AFU Clerical Matching operation.

7.7 Clerical Geocoding and Clerical Residence Status Coding

After automated geocoding and residence status coding, the following types of cases will be reviewed during CGC and RSC:

- Cases with ungeocoded inmover or alternate addresses - Any inmover or alternate address that is not geocoded to the block level by automated geocoding will be reviewed by the technicians. They will have access to census maps for the entire country, the MAF browser, the online ZIP+4 directory, and the Internet in order to attempt to geocode the address. They will also have access to other information collected in the PI such as landmarks, cross streets, names of other household members, and names of nearby neighbors.
- PI cases with a “review” residence status code - In some cases, the computer cannot assign a residence status code to a person rostered in the PI, but notes collected during the

²³PerMaRCS is a software system designed by Gunnison Consulting Group Inc. that allows clerical matchers to review records and assign match codes, residence status codes, and sufficiency status codes.

²⁴ See Attachment 3 for more information on the sufficiency and inmover address codes.

PI may help resolve the status. In these cases, a technician will review the data collected in the PI and will assign a residence status code.

The analysts will perform a review of selected technicians' work. They will also review specific records referred to them by the technicians. We estimate that 16.7 percent of addresses will be subject to clerical geocoding verification with an AOQL of 7.7 percent. For clerical RSC, approximately 16.7 percent of persons will be subject to verification with an AOQL of 7.4 percent.

7.8 Before Followup Technician and Analyst

Next, during Before Followup Technician (BFT), technicians will review the results of Person Computer Matching. In particular, technicians attempt to match PI records to the census, find duplicates in the PI, and find E-sample duplicates.

The technicians will review each E-sample case with insufficient information for followup to attempt to update the record with information from other census data (including electronic images of census forms). An example of this type of updating would be adding the last name from the first person listed on a census form to other household members on the form whose last names are missing. PI cases with incomplete names or characteristics will be updated using other PI data.

Last, as part of Before Followup Analyst (BFA), analysts perform a 100-percent dependent review of selected technicians' work and resolve any cases that the technicians could not resolve. Technicians who qualify based on the quality of their previous work will have their verification percentage lowered. The analysts will perform a full review of block clusters selected for the matching quality control sample. They will also review specific person records sent by the technicians in BFT. It is expected that about 16.7 percent of all block clusters requiring person matching will be subject to this quality control operation, with an expected AOQL of 4.1 percent.

After this initial matching operation, all E-sample nonmatches, selected PI nonmatches, and all possible matches and duplicates are followed up in the field via the PFU operation.

7.9 Person Followup

Person Followup (PFU) cases are identified after the Person BFU Clerical Matching operation is completed. The PFU interview attempts to collect additional information needed to establish Census Day residence for P-sample cases. It also attempts to resolve enumeration status for nonmatched E-sample cases. Persons coded as having insufficient information for followup are not sent to PFU. Table 5 shows the types of cases sent to PFU (Shoemaker, 2008d)

Table 5: Cases Sent to CCM Person Followup

Possible matches (P or E sample)
P-sample partial household nonmatches with a proxy respondent in the PI
P-sample whole household nonmatches with a proxy respondent in the PI and either no people in the matching census unit or no matching census unit
P-sample whole-household nonmatches in clusters with a high rate of P-sample nonmatches (more than 45 percent) will be flagged for a geocoding check
P-sample whole-household nonmatches where the PI interviewer changed the sample address will be flagged for a geocoding check
Inmovers with ungeocoded inmover addresses (regardless of match status)
Conflicting households (nonmovers or outmovers in the P sample with completely different rosters in the census and PI)
Unclassified P- or E-sample persons who are not duplicates, regardless of match status
P-sample people with a proxy respondent in the PI who are identified as match or possible match in the nationwide computer matching (followup at sample address)
Census people who match or possibly match to P-sample people in the nationwide computer matching (followup by telephone at the address outside the search area done from a Census Telephone Center)
E-sample nonmatches ²⁵
E-sample whole household nonmatches whose addresses were not included in Initial HU Matching
E-sample persons who are a duplicate or possible duplicate to a person identified in the nationwide computer matching if there are no PI results or if there was a proxy respondent in the PI (followup at sample address)
People identified in the nationwide computer matching who are duplicates or possible duplicates to E-sample people (followup by telephone at the address outside the search area done from a Census Telephone Center)
Any case sent to PFU by an analyst
P-sample possible duplicates within the sample block cluster
E-sample possible duplicates within the sample block cluster/surrounding blocks (at least one of the pair is an E-sample person)
Inmovers who do not match at their Census Day inmover address, with a PI proxy respondent

Note that a case can be sent to PFU for multiple reasons. The PFU data collection methods will include establishing where the person should have been counted in the census (their Census Day residence) and collecting information on alternate locations and addresses where the person could have been counted on Census Day.

The PFU interview will be conducted using a docuprinted paper questionnaire. This method allows the PFU form to be tailored to collect only the information needed for each person. The paper questionnaire will be used for final coding of the Person Matching results. After use in Person AFU Clerical Matching, the forms will be keyed to capture data needed for estimation and evaluation activities.

The PFU questions are geared towards collecting addresses at any of the following places where a person might have resided:

- Census Day address other than the sample address
- college address
- military service address
- residence for a job

²⁵ Persons enumerated by the census in Transitory Locations are not eligible for PFU, even if they are selected to be in the E sample. The chances that the same person will be at that location nearly a year after Census Day are very small.

- relative's address where the person may stay, including joint custody
- seasonal home address
- any other place stayed often
- other GQs (nursing home, correctional facility, group home, etc.)

In order to assign a residence status code, the PFU interview then asks for dates spent at each residence and includes some semi-scripted probes to help matching staff determine at which address the person should be counted on Census Day according to Census Residence Rules.

In addition, a centralized telephone operation at one of the Census Telephone Centers (CTC) will followup on possible "long distance" duplicates identified in the nationwide computer match. The part of the link that is located outside of the CCM sample and surrounding blocks will be followed up by telephone. The part of the link located inside the search area will be followed up in the field as part of PFU if it is eligible for followup. CTC supervisors will monitor the phone calls for quality control purposes.

The PFU field operation is subject to several quality checks. Crew leaders observe the interviewers to make sure that procedures are being followed correctly. They also conduct edits of the completed PFU forms. As was done for the PI, a dependent quality control interview will check for falsification. A different interviewer will visit the followup address or make contact by telephone and conduct the reinterview with the same person who responded to the PFU interview to confirm that the PFU interview was done. The purpose of the reinterview is to ensure that the original interviewer contacted the PFU case. About 10 percent of each interviewer's completed eligible cases will be subject to reinterview.

7.10 Person After Followup Clerical Matching

Following the PFU interview, Person AFU Clerical Matching operations are performed. The Person AFU Clerical Matching consists of four stages: After Followup Technician (AFT), After Followup Analyst (AFA), After Followup Review (AFR), and After Followup Outlier Review (AFO).

As PFU forms are returned from interviewing, they are assembled into batches to be reviewed by the clerical matching staff. First, these batches are reviewed by technicians in AFT. Then, during AFA, analysts perform a review of the technicians' work selected for quality control purposes. The analysts also resolve any cases that the technicians could not resolve. This process is similar to the process for all other CCM clerical matching operations. About 16.7 percent of all PFU forms will be reviewed for quality control purposes. The target AOQL is 9.2 percent.

During AFT and AFA, the PFU interview results are clerically reviewed and coded according to a pre-determined set of matching codes. Technicians and analysts are allowed to make corrections to previous coding assignments if errors are discovered from the Person BFU Clerical Matching operations. These match codes provide the basic information needed for measuring both net coverage error and the component errors. This also includes clerical geocoding of any additional respondent-provided addresses.

All CCM sample block clusters are eligible for AFR and AFO. Clusters are flagged for AFR by technicians or analysts in the AFT or AFA stage when certain conditions make it necessary to review the cluster as a whole instead of the batched forms. AFO is an analyst-only stage that targets the block clusters that may need further review. Analysts will review the block cluster and write journals to explain any problems with the block cluster. Analysts will also recode any cases that may be incorrectly coded. The journals are useful during analysis when particular block clusters stand out. Block clusters are flagged for outlier review by headquarters staff or by having a high outlier priority score. The outlier priority score is computed as the weighted sum of cases that do not match in the net error context in the P sample and cases in the E sample that do not link back to the P sample or are considered erroneous enumerations for net error.

Person AFU Clerical Matching is the final CCM person operation. An output file with the results from the PI, person matching, and PFU operations will be made available for the CCM Estimation activities (see Section 9). Persons that remain unresolved will be handled through statistical techniques for missing data.

8.0 Final Housing Unit Operations

The CCM Final HU operations consist of the Final HU Computer Processing, Final HU BFU Clerical Matching, Final HU Followup, and Final HU AFU Clerical Matching operations. As a whole, these four operations work together to: identify geocoding errors and any remaining housing unit duplicates, update pre-existing files, and resolve match, enumeration status, and housing unit status for all P- and E-sample housing units not resolved in prior operations.

8.1 Final Housing Unit Computer Processing

In the Final HU Computer Processing, information from all previous stages of computer and clerical matching from both housing unit and person procedures, coupled with final census data, is used to update housing unit records. Using the person matching results, the Computer Processing assigns geocoding error codes to P- and E-sample units where the people are coded as such, and assigns Census Day housing unit status to unresolved housing units based upon specific PI results. These PI results include vacant housing units and records that were not housing units on Census Day.

In addition, Final HU Computer Processing identifies units that will need to go to Final HU BFU Clerical Matching. Computer Processing will assign a work flag to the following types of units (Shoemaker, 2008e):

- P-sample links to census UCM deletes²⁶
- census UCM adds²⁷ that are in the sample block cluster (including subsampled out)
- census UCM adds that are in the surrounding blocks to 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

²⁶ A census UCM delete is a housing unit on the preliminary census address list used for Initial HU Matching, but NOT on the CUF.

²⁷ A census UCM add is a housing unit NOT on the preliminary census address list, but is on the CUF.

- E-sample units that linked to a unit that is no longer in the P sample
- any P- or E-sample unit (not covered above) for which Housing Unit Computer or Clerical Matching has not yet been conducted²⁸

Lastly, Computer Processing identifies CCM and census units, including census GQs, with links to census units not on the CUF, and then un-links those units. Once the data files have been updated and the units flagged, the next process can begin, the Final HU BFU Clerical Matching.

8.2 Final Housing Unit Before Followup Clerical Matching

In the Final HU BFU Clerical Matching operation, files containing the entire block clusters including the flagged housing units from Computer Processing are made available for matchers to view. Matching staff will attempt to match by address, housing unit description, or map spot location, all P-sample units and census units that were flagged earlier. Those units that remain unresolved will be flagged for FHUFU.

In addition to identifying matches between CCM and census units, matching staff will attempt 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

Also, both technicians and analysts are allowed to change match codes previously assigned if errors are uncovered, and may attempt to resolve the case without a field visit by Internet research.

The Final HU BFU (and AFU) Clerical Matching procedures are also subject to a quality control process in order to assure accurate results. The quality control process is similar to that used for other CCM Housing Unit Clerical Matching operations previously described, with approximately 16.7 percent of block clusters subject to quality control and a target AOQL of 4.0 percent.

8.3 Final Housing Unit Followup

The FHUFU is a field operation performed by CCM field staff. In making assignments, care is taken to observe the independence rules followed for all CCM operations. Staff will use a preprinted paper form and paper maps with map spots from the CCM and the census in order to locate flagged housing units. Interviewers may annotate the CCM maps if they feel it will help explain the situation.

The questions on the FHUFU forms will be tailored for each specific case given its situation. The following information will be collected:

²⁸ All P-sample housing units should have been processed through Initial HU Computer and/or Clerical Matching by point. They are included here to handle unexpected problems that might prevent a cluster from going through the Initial HU Matching system.

- census collection block number
- information to determine if there was a housing unit at the address on Census Day. If not, why not?
- information to determine whether or not units identified as possible matches really are the same
- information to determine whether or not units identified as possible duplicates really are the same

The questions will be the same as those on the CCM IHUFU form for similar types of followup situations, except for changes needed due to the different reference dates for the initial and final followup interviews.²⁹

Interviewers will use the reference list³⁰ provided during the FHUFU to determine if any units sent to followup match other units on the reference list. All of the information collected during followup will be used to assist in determining the status of followed up housing units as of Census Day. As with all other field operations, FHUFU is subject to a quality control procedure very similar to that used for IHUFU, with about 13.5 percent of FHUFU forms subject to quality control and a target AOQL of 6.7 percent.

8.4 Final Housing Unit After Followup Clerical Matching

The Final HU AFU Clerical Matching operation will have multiple sources of information to help resolve housing unit status and match status such as:

- data used for CCM Final HU BFU Clerical Matching
- completed FHUFU and IHUFU forms
- completed PFU forms
- CCM Final HU BFU Clerical Matching results
- CCM Person AFU Clerical Matching results
- paper copies of CCM maps used/annotated during FHUFU

Utilizing this wealth of data, matching staff will attempt to assign a match/enumeration status code to each case sent to FHUFU. An output file with the results from the CCM housing unit operations will be made available for the CCM Estimation activities. Housing units that remain unresolved will be handled using statistical techniques for missing data.

9.0 Estimation

The CCM Estimation process consists of several operations, which will ultimately lead to the production of estimates of coverage errors for both housing units and persons in housing units (hereafter referred to as “persons”). This includes estimates of net coverage error as well as estimates of the component errors— omissions and erroneous enumerations.

²⁹ If any serious problems are encountered with the IHUFU form, the FHUFU form may be changed.

³⁰ The reference list is a list of all CCM and census housing units in the CCM sample block cluster and its surrounding blocks.

It should be noted that GQ facilities and persons residing in those facilities are not within the scope of the CCM Program. The estimation of coverage error components is new for the 2010 CCM Program. Previously, the Census Bureau released estimates of erroneous enumerations and omissions only at a national level. Since, these estimates were based on available data from the matching operations and several simplifying assumptions, they were subject to some deficiencies. To address some of the deficiencies in previous component error estimates, the 2010 CCM survey design includes expanded interviewing, matching and followup operations to provide the necessary information for both net error and component error estimation.

Like the 1990 Post Enumeration Survey and the 2000 A.C.E., the 2010 CCM will be estimating net coverage error by using DSE to generate the population estimates of housing units and persons in housing units. We will use logistic regression modeling instead of the traditional post-stratification to estimate the parameters in the DSE formula. Logistic regression modeling will also be used in some of the missing data operations; these will be described in more detail later.

Since estimation of coverage error components is new for the 2010 CCM, we will be using a more straightforward estimator; a basic inflation of ratio-adjusted survey weights. The ratio adjustment includes two stages to reduce the variance of the estimates. Also, the adjustment ensures that the estimates of correct and erroneous enumerations add to the appropriate number of data-defined census enumerations.

The remaining sections of this document summarize the more detailed description of the CCM Estimation methodology, which can be found in Mule (2008). Section 9.1 covers missing data for P-sample persons. Section 9.2 covers missing data for E-sample persons. Section 9.3 covers the estimate of net coverage error for persons. Section 9.4 covers the components of coverage error for persons. Sections 9.5 to 9.8 mirror sections 9.1 to 9.4, with housing units rather than persons as the focus.

Although not specifically addressed in the following sections, a weight trimming procedure is applied after the missing data operations have been performed, but before the final coverage estimates are created. Trimming the weights reduces the likelihood that a small number of block clusters will have a large influence on the coverage estimates and their standard errors. The process we are planning to use is very similar to the one used for the 2000 A.C.E. The cutoff values are different for housing units and for persons. The procedure used for the A.C.E. is documented in Appendix C of U.S. Census Bureau (2004). Attachments 5 and 6 outline the estimation operations for persons in housing units and housing units, respectively.

9.1 Missing Data for P-Sample Persons

There are four adjustments designed to account for missing data for persons in the P sample: imputing missing demographic characteristics, making a household-level noninterview

adjustment for the CCM PI, imputing the probability that a person should be included in the P sample, and imputing a match status.

9.1.1 Imputing Missing Demographic Characteristics for Persons

Production of CCM coverage estimates requires data on age, sex, tenure (owner versus non-owner), race and Hispanic origin to classify respondents by these important demographic characteristics so they will be imputed whenever the data was not collected. Our current plan is impute these characteristics in a manner that is consistent with what the census is doing to the extent possible. For cases with missing age, we will impute a single year of age. This was done for A.C.E. Revision II but not for the original A.C.E. production estimates. The A.C.E. used a hot-deck procedure to assign missing tenure, race, and Hispanic origin. The current plan is for the CCM to utilize the Census Edit and Imputation programs with necessary changes to account for the design differences between CCM and census.

9.1.2 Household-Level Noninterview Adjustment for CCM Person Interview

The CCM accounts for a potential bias being introduced because of households that were not interviewed in the PI. These households were not interviewed because they could not be contacted or because the interview was refused. The units with 'missing' interviews have sample weights associated with each. The noninterview adjustment procedure accounts for these cases by assigning a weight of zero to noninterviews and distributing the weight of the noninterviewed cases uniformly among interviewed cases that possess similar qualities. The distribution will be done for housing units within similar types of structures in the block cluster unless the ratio of noninterviewed to interviewed housing units is too large. If so, then the excess weighted total above the maximum allowable ratio will be distributed to other housing units in the block cluster. There is a contingency plan in place if the amount of weight distribution is too large even at the block cluster level. For the 2000 A.C.E., the maximum ratio was 2.

The CCM will use the same adjustment procedure as used for the A.C.E. except for a different approach for movers. Only a noninterview adjustment for Interview Day will be done. For 2010, the CCM PI will interview persons who have moved into the sample housing unit and certain persons who have moved away since Census Day. In 2000, the A.C.E. interviewed both in-movers and out-movers. Therefore, A.C.E. implemented two noninterview adjustments since they were collecting information for two points in time: one for Interview Day and a second for Census Day.

9.1.3 Imputing the Probability of P-Sample Inclusion for Persons

The CCM interview needs to determine if a person should be included in the P sample. The P sample of persons includes residents of the P-sample housing unit on Interview Day, and out-movers from the P-sample housing unit who were residents on Census Day and who have no chance of selection in the P sample based on their Interview Day residence. One example is a person who resided in the P-sample housing unit on Census Day and now resides in a group quarters on Interview Day.

For out-movers, only those persons who have been resolved as residing on Interview Day in a living quarters which has no probability of being selected in the P sample will be included in the P sample. This decision is based on the tradeoff between having representation in our sample for

this segment of the population versus introducing an error of including people who are eligible to be included in the P sample in another housing unit.

For the remaining cases that are eligible to be included in the P sample, the CCM has implemented similar changes for this imputation which are similar to the ones being developed for the enumeration status imputation. The first is using logistic regression modeling instead of Imputation Cell Estimation. The second involves expanding this operation to account for the potential error in overestimating the number of cases belonging to the P sample. This is based on utilizing information on the cases that link outside the net error search area.

For nonmovers, we have determined that their Census Day residence and their Interview Day residence are the same. The finding of a link to another census enumeration outside of the search area raises the possibility that this case may not have fully reported the information about their Census Day residence. Since we have measured both the Census Day and Interview Day residence to be the same, we are assuming that we may have made an error about the Interview Day status and may be including some wrongly in the P sample. For in-movers, we are making a different assumption. We are assuming that since we found a link to another census enumeration, the person may not have fully reported their information, this raises a possible question about whether we have captured their Interview Day status.

We will assign a probability of P-sample inclusion status for all sample cases. Resolved cases that should be included in the P sample will primarily receive a probability 1. Resolved cases that should not be included will receive a probability of 0. The cases with unresolved status need to have a probability assigned to them. We will use logistic regression to develop one or more models to account for cases with this unresolved status. We will implement one or more logistic regression models on the resolved cases to determine the regression coefficients.

Using the coefficients from the logistic regression(s) described in the previous paragraph and the values of the independent variables of the unresolved cases, we can impute the probability of being in the P sample. Since the overstatement of the P sample total was an error for the A.C.E., the CCM Estimation will closely examine and review how the imputation handles these cases. The details of this review are being determined.

9.1.4 Imputing the Probability of Match Status for Persons

The CCM needs to determine if the P-sample person matched to a census enumeration in the net error search area³¹. This involves accounting for P-sample cases that may have an unresolved match status or an unresolved mover status. A record may have either or both of these statuses being unresolved. This section lays out the methodology to account for both statuses and how the final match probability for net error estimation is assigned.

We will assign a probability of match status for all sample cases. Resolved match status cases will receive a probability 1. Resolved non-match status cases will receive a probability of 0. The cases with unresolved status need to have a probability assigned to them.

³¹ The net error search area is typically the CCM sample block cluster and one ring of surrounding blocks.

Since both the match status and mover status can be unresolved, we will assign an overall match probability for unresolved cases based on the following conditional probability of mover status (nonmover or inmover).

We first need to account for unresolved mover status. Resolved mover status cases will either be nonmovers or inmovers. Using the resolved cases, we will determine the weighted proportion of cases that were nonmovers and inmovers. This will be done by forming cells using some of the covariates in the logistic regression models like BFU groupings. This will allow the two mover probabilities to be assigned to each case: the probability of being a nonmover and the probability of being an inmover. The weighted proportions will be assigned as these probabilities for the unresolved mover status cases.

We will then assign the probabilities of a case being a match given that it is either a nonmover or an inmover. When these conditional probabilities are unresolved, we will use logistic regression methods to impute them. These models and the general prediction equation will be used to determine the probability of cases matching given that the record is either a nonmover or an inmover.

The predicted match rate for each unresolved match status case given that it is a nonmover is then obtained. If the case is a nonmover then the correct search area is the block cluster and ring of surrounding blocks. We first use the clerical matching information to see if they were able to match the case in the block cluster search area. If so then we will assign this conditional probability of a match for this case equal to 1. The clerical matching information may indicate that the case was a nonmatch in the block cluster search area. If so then we will obtain this conditional probability of a match equal to 0. If the matching results are unresolved, then we will assign the conditional probability of a match by using logistic regression. If mover status was used as a covariate in the logistic regressions and the P-sample case has an unresolved mover status, then for this conditional probability assignment, the P-sample case will be treated as being a nonmover for this prediction.

The predicted match rate for each unresolved P-sample case given that it is a inmover is obtained next. One difference from the conditional match probability given that the case is a nonmover is that the correct search area for inmovers needs to be identified based on interview information. The clerical results may indicate that a case links to a census enumeration outside the block cluster search area but we cannot assume that it is the correct search area. For these cases the conditional match prediction will be assigned using logistic regression. If mover status was used as a covariate in the logistic regressions and the P-sample case has an unresolved mover status then for this conditional probability assignment, the P-sample case will be treated as an inmover in this prediction.

After a) the probabilities of the case being either a nonmover or inmover and b) the conditional match probabilities of nonmovers and inmovers are estimated, we can estimate the probability of the unresolved case being a match.

9.2 Missing Data for E-Sample Persons

There are two missing data procedures needed for persons in the E sample: a procedure for dealing with missing demographic characteristics and one for imputing enumeration status (correct or erroneous). However, imputation of enumeration status needs to be considered separately depending on whether we are estimating net or component coverage error. Each of these processes will be discussed in more detail in the following sections.

9.2.1 Missing Demographic Characteristics for E-Sample Persons

For persons in the E sample with missing demographic characteristics, we will not develop any special imputation procedures. We will match those records to the Census Edited File and obtain the imputed characteristics from that file based on the Census Edit and Imputation methodology.

9.2.2 Imputing Missing Enumeration Status for Persons in the E Sample (Net Error)

The CCM needs to determine if the census enumeration in the E sample was correctly enumerated in the net error search area. The probability of correct enumeration will be assigned for all E-sample cases. The cases with unresolved status will have a probability imputed for them. The imputation procedure will attempt to minimize the potential measurement error without introducing other biases.

The 2010 CCM Program has changed the imputation of enumeration status in two major ways. The first involves using a logistic regression model instead of imputation cell estimation to generate the predicted probability of unresolved cases being correct enumerations. The independent variables in the logistic regression models used to predict correct enumeration probability will include matching and CCM interview information that has been identified as good discriminators of enumeration status that are only available for cases in the E sample and not available for the entire census. The models will include, at a minimum, the main effects of the independent variables used in the regression models to determine the predicted probabilities of being data-defined, a correct enumeration, or a match in the overall dual system estimation formula, as noted in Section 9.3.

The second involves addressing a major error in the A.C.E. of overstating the number of correct enumerations by coding cases that have been linked to an enumeration outside of the search area as being unresolved. During clerical matching, the CCM will utilize the nationwide computer matching results to identify E-sample cases who have a possible other Census Day residence. This additional matching may identify people who may have not fully described their living situation during the CCM PI or PFU interview. The probability of correct enumeration will be adjusted by a factor that accounts for the number of duplicate links. It will also adjust the correct enumeration probability for the number of duplicate links between that person and other persons enumerated in the block cluster search area that were not selected for the E sample.

9.2.3 Imputing Missing Enumeration Status for Persons in the E Sample (Component Error)

A strict definition of correct enumeration is used to employ the DSE methodology for net error in order to minimize matching error. This definition is not appropriate for estimating coverage error components, however, since it results in inflated estimates of erroneous enumerations. Therefore, some cases treated as erroneous enumerations for net error estimation are correct enumerations for component error estimation. One example of this is that an enumeration may be treated as erroneous for net error because it was enumerated in the wrong location but, if the person was enumerated once and only once, then that enumeration would be correct for national estimates of components.

Another example is the treatment of cases with insufficient information for net error. If they are data-defined, but don't have adequate information for CCM followup, they are treated as resolved erroneous enumerations for net error. However, for component error estimation, the CCM Program has expanded its matching operations to try to determine the enumeration status of these cases. If this matching effort is successful, some of the cases treated as erroneous enumerations for net error will be correct enumerations for estimating components.

For some cases, the additional matching will not be able to resolve the enumeration status. Because of the lack of information collected (that being usually name) the case is unable to go to PFU to be resolved and is thus unresolved for component estimation. The imputation scheme for component estimation needs to be able to account for these unresolved cases.

As part of estimating component error, probabilities of various outcomes not needed to implement the DSE net error methodology need to be assigned. These outcomes include various reasons for the case being classified as an erroneous enumeration. Therefore, the CCM Program will be implementing a new missing data adjustment for enumeration status for component estimation.

Since we are doing components for the first time, we will be using an Imputation Cell Estimation methodology to assign the probabilities of the various statuses for the unresolved cases. The A.C.E. used this methodology for the missing data adjustments of unresolved enumeration statuses for net error estimation in 2000.

The CCM Estimation staff is researching what census and CCM data can be utilized to form the best set of imputation cells for 2010. The staff is considering whether data provided during PFU can be utilized to provide more discriminating cells for the imputation. The weighted averages of the correct and erroneous enumerations (overall by type based on national definition and geographical definition) will be estimated using the resolved cases and those weighted averages will be assigned to the unresolved E-sample cases.

For component estimation, we will assign probabilities for the following unresolved statuses:

- correct enumeration nationally
- erroneous enumeration nationally

- erroneous enumeration who should not have been enumerated at all in a housing unit
- erroneous enumeration due to duplication
 - due to duplication to people in other housing units by geographic distances
 - due to duplication to people in GQs by groupings of facilities
- erroneous enumeration in the wrong location
 - right county, right state (but outside the net error search area)
 - wrong county, right state
 - wrong state

The CCM Program will not be evaluating whether census imputations are correct or erroneous. The total number of census imputations will be reported as part of the component estimation, and no missing data adjustment for enumeration status will be done for these cases.

9.3 Person Net Coverage Error

There is a long history of using DSE in measuring coverage errors in a census. The DSE can be expressed as:

$$DSE = \sum_{j \in \mathcal{U}} \frac{\bar{d}(j)}{\bar{m}(j)} \times CB_j$$

The predicted data-defined, correct enumeration, and match probabilities (π_{dd} , π_{ce} , π_m , respectively) will be obtained through logistic regression modeling. The correlation bias adjustment factor (CB_j) is computed using sex ratios from the Census Bureau's Demographic Analysis program. These need to be adjusted to account for differences between the CCM target population and the census universe; namely removing persons in GQs and in remote Alaska.

We will then estimate net error by comparing the estimate of the true population (from the DSE) to the census count using the following formula:

$$NetError = TruePopulation - Census$$

9.4 Person Components of Coverage Error

For persons in housing units, separate estimates of correct enumerations, erroneous enumerations, omissions, and census imputations will be formed at the national level and for various subpopulations (or estimation domains) using data from the E sample. The requirements for an enumeration to be considered correct for component error estimation are less stringent than those treated as correct for net error. We will tally and report the total number of census whole-person imputations overall and for various sub-categories.

9.4.1 Two-Stage Ratio Adjustment for E-Sample Persons

Estimates of correct and erroneous enumerations will be formed by inflating ratio-adjusted survey weights attached to person records from the CCM E sample in the particular estimation domain. The ratio adjustment consists of two stages.

In the first stage, a set of pre-defined cells is formed and the sum of the sampling weights for the E-sample cases in each cell is ratio adjusted so they equal the census totals for those cells. These cells can be based on race/ethnicity, tenure, age/gender groupings or other demographic variables and the census totals are the number of data defined census person records in each cell.

Depending on sample sizes and the size of the adjustment factors, some collapsing of the original cells may be necessary.

The second stage of the ratio adjustment is done when generating an estimate for a particular domain. In each case, the sum of the first-stage ratio adjusted sampling weights for cases in that domain will be ratio adjusted to the census total for the domain. As in the case for the first-stage adjustment, the data-defined census person count will be used as the census count for person estimates.

9.4.2 Erroneous and Correct Enumerations for Persons

For component estimation, the CCM will report the total number of erroneous enumerations. The CCM Program will also estimate erroneous enumerations by type:

- Persons who should not have been enumerated at all
- Duplicate person enumerations

The CCM is using the following definition of being a correct person enumeration when evaluating the two coverage universes in the 2010 Census:

The enumeration is considered to be correct if the record corresponds to a person that should have been included anywhere in the coverage universe³². If such a person was included multiple times, one of the enumerations will be considered correct and the other enumerations will be erroneous.

Based on our implementation of correct enumeration for components of census error, a person will be considered correctly enumerated by being included in a housing unit anywhere in the U.S.

9.4.3 Person Omissions and Omissions by Missed Housing Unit

An overall estimate of the number of persons omitted by the census will be obtained by adding the net error estimate (obtained from DSE) to the total estimated number of erroneous enumerations:

³² The CCM coverage universe includes housing units and person in housing units. It does not include GQs, or Remote Alaska Enumeration areas.

$$\text{Omissions} = \text{net error} + \text{EEs}$$

For persons, separate estimates of omissions will be formed by whether or not the housing unit was enumerated in the census. The estimate of person omissions by housing unit inclusion status is estimated by the proportion of resolved nonmover person nonmatches that were in housing units that either a) matched or b) did not match.

9.5 Missing Data for P-Sample Housing Units

There are three missing data operations for housing units in the P sample: imputing housing unit characteristics, imputing housing unit status, and imputing housing unit match status. Each of these will be discussed in more detail in the sections that follow.

9.5.1 Imputing Missing Characteristics for P-Sample Housing Units

Production of CCM housing unit coverage estimates requires data on occupancy status, tenure (owner versus non-owner), and race/ethnicity domain of the householder to classify housing units by these important characteristics. So, they will be imputed whenever the data was not collected. The details of the imputation process are still being developed, but it will be designed to be as consistent as possible with the Census Edit and Imputation System.

9.5.2 Imputing Missing Housing Unit Status for P-Sample Housing Units

The CCM Housing Unit Matching operation needs to determine if a P-sample housing unit was a housing unit on Census Day. For some units, this will not be able to be determined and the housing unit status will be unresolved. We will assign a probability of housing unit status for all sample cases. Resolved housing units will receive a probability of 1. Resolved cases determined not to be housing units will receive a probability of 0. The cases with unresolved status need to have a probability assigned to them.

We will use logistic regression to develop one or more models to account for cases with this unresolved status. We will implement one or more logistic regression models on the resolved cases to determine the regression coefficients. Using these coefficients and the data of the unresolved cases we can then estimate a probability of being a housing unit for these unresolved cases.

9.5.3 Imputing Missing Match Status for P-Sample Housing Units

The CCM Housing Unit Matching operation needs to determine if the housing unit matched a census enumeration in the net error search area. This process is simpler for housing units than for persons, because mover status does not need to be considered. We will assign a probability of match status for all sample cases. Resolved match status cases will receive a probability 1. Resolved nonmatch status cases will receive a probability of 0. The cases with unresolved status need to have a probability assigned to them.

We will use logistic regression to develop one or more models to account for cases with unresolved match status. We will implement one (or more if necessary) logistic regression models on the resolved cases to determine the regression coefficients. Using these coefficients and the data of the unresolved cases we can then estimate a probability of match for the unresolved cases.

9.6 Missing Data for E-Sample Housing Units

There are two missing data procedures needed for housing units in the E sample: a procedure for dealing with missing housing unit characteristics and one for imputing enumeration status (correct or erroneous). Imputation of enumeration status needs to be considered separately depending on whether we are estimating net or component coverage error. Each of these processes will be discussed in more detail in the following sections.

9.6.1 Missing Characteristics for E-Sample Housing Units

Production of CCM housing unit coverage estimates requires data on occupancy status, tenure (owner versus non-owner), and race/ethnicity domain of the householder to classify housing units by these important characteristics. So, they will be imputed whenever the data was not collected. For cases in the E sample with missing housing unit characteristics, we will not do any special imputation procedures. Instead, we will use the CEF and obtain the imputed characteristics from there.

9.6.2 Imputing Missing Enumeration Status for E-Sample Housing Units (Net Error)

The probability of being a correct enumeration for net error will be determined using a two-step procedure. The first will assign an initial correct enumeration probability using logistic regression. The second will adjust that initial probability using a factor that accounts for the number of duplicate links between that housing unit and other E-sample housing units inside the sample block cluster search area. It will also adjust the correct enumeration probability for the number of duplicate links between that housing unit and other housing units enumerated in the block cluster search area that were not selected for the E sample.

9.6.3 Imputing Missing Enumeration Status for E-Sample Housing Units (Component Error)

The CCM Program will be using a less stringent definition of correct enumeration status for component error estimation than used for net error estimation. With this definition, there will be some cases that were determined to be erroneous enumerations for net error estimation that will be unresolved enumerations for component estimation.

One example of the difference is that an enumeration may be erroneous for net error estimation because it was a geocoding error. For component estimation, a housing unit is correct if it was included once and only once in the census. These geocoding errors have been erroneously assigned to the sample block cluster. For our estimation, the missing data model will assign

probabilities to these misgeocoded census housing units to reflect their being correct or erroneous for component estimation.

9.7 Housing Unit Net Coverage Error

The estimate for housing units does not include the correlation bias adjustment, and, since the data-defined concept does not apply to housing units, the associated predicted probability of being data-defined is excluded. The DSE for housing units can be expressed as:

$$DSE = \sum_{j \in m(j)} \overline{ce(j)}$$

The predicted correct enumeration and match probabilities (π_{ce} , π_m , respectively) will be obtained through logistic regression modeling.

We will then estimate net error by comparing the estimate of the true population (from the dual system estimator) to the census housing unit count using the following formula:

$$NetError = TruePopulation - Census$$

9.8 Housing Unit Components of Error

The process for computing component error estimates for housing units is very similar to the one used for persons, including a two-stage ratio adjustment as described in Section 9.4.1. The ratio adjustment cells can be based on region, occupancy status or other variables and the cell totals are the corresponding census HU counts. Depending on sample sizes and the size of the adjustment factors, some collapsing of the original cells may be necessary.

For housing units, separate estimates of correct enumerations, erroneous enumerations, and omissions will be computed. The erroneous enumeration estimate will be broken down into two parts: structures enumerated in the census as housing units, but were not housing units; and housing units enumerated more than once (duplicates).

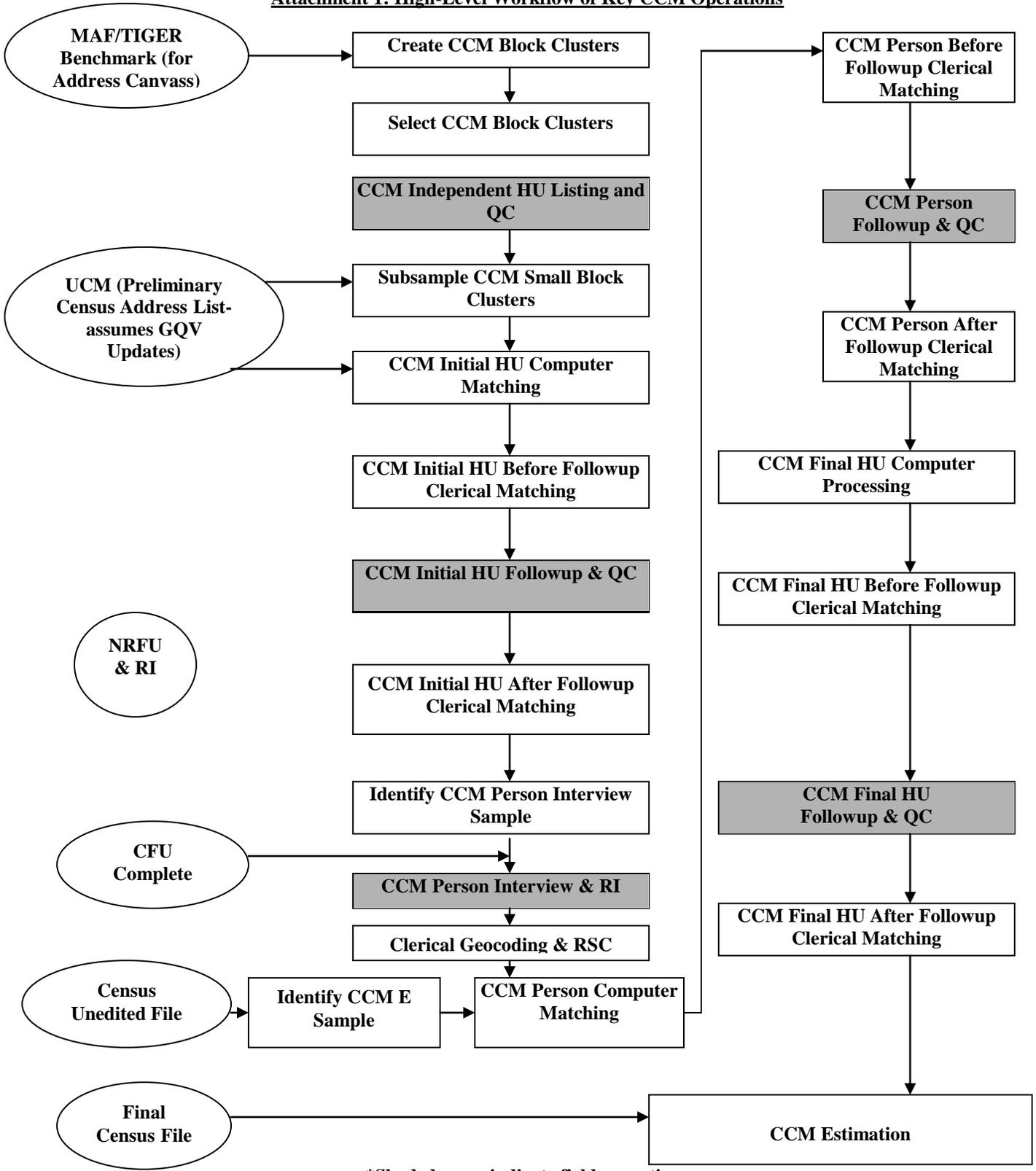
As with person omissions, the estimated number of housing units omitted from the census will be obtained by adding the estimate of net error (from DSE) to the estimate of erroneous enumerations.

10.0 References

- Davis, P. (2008) "2010 Census Coverage Measurement: Block Cluster Sample for Independent Listing Software Requirements Specification." DSSD 2010 Census Coverage Measurement Memorandum Series #2010-C-05.
- Fay, R.E., Passel, J.S., and Robinson, J.G. with assistance from C.D. Cowan. (1988) *The Coverage of Population in the 1980 Census*, 1980 Census of Population and Housing, Evaluation and Research Reports PHC80-E4. Washington, D.C.: U.S. Department of Commerce.
- Hogan, H. (1993). "The 1990 Post-Enumeration Survey: Operations and Results." *Journal of the American Statistical Association*, Vol. 88, No. 423 (Sep., 1993), 1047-1060.
- Kostanich, D. (2006). "2010 Census Coverage Measurement – Summary of Changes." DSSD Census Coverage Measurement Memorandum Series #A-xx, draft dated: November 30, 2006.
- Mule, T. (2002). "A.C.E. Revision II Results: Further Study of Person Duplication," DSSD Revised A.C.E. Memorandum Series PP-51.
- Mule, T. (2008). "2010 Census Coverage Measurement Estimation Methodology." DSSD 2010 Census Coverage Measurement Memorandum Series #2010-E-xx--DRAFT
- Mule, T. et al (2008). "2010 Census Coverage Measurement Sample Design." DSSD Census Coverage Measurement Memorandum Series #2010-C-xx--DRAFT
- Shoemaker, H. (2008a). "High-Level Design Requirements for 2010 Census Coverage Measurement Independent Listing Operation." DSSD 2010 Census Coverage Measurement Memorandum Series #2010-B-04.
- Shoemaker, H. (2008b). "High-Level Design Requirements for 2010 Census Coverage Measurement Initial Housing Unit Matching and Followup." DSSD 2010 Census Coverage Measurement Memorandum Series #2010-B-05.
- Shoemaker, H. (2008c). "High-Level Requirements for the 2010 Census Coverage Measurement Person Interview and Reinterview." DSSD 2010 Census Coverage Measurement Memorandum Series #2010-B-06.
- Shoemaker, H. (2008d). "High-Level Design Requirements for 2010 Census Coverage Measurement Person Matching and Followup." DSSD 2010 Census Coverage Measurement Memorandum Series #2010-B-xx-DRAFT.

- Shoemaker, H. (2008e). "High-Level Design Requirements for 2010 Census Coverage Measurement Final Housing Unit Computer Processing, Matching, and Followup." DSSD 2010 Census Coverage Measurement Memorandum Series #2010-B-xx-DRAFT.
- Singh, R. (2005). "2010 Census Coverage Measurement – Updated Plans." DSSD 2010 Census Coverage Measurement Memorandum Series #A-06.
- U.S. Census Bureau (2004). "Accuracy and Coverage Evaluation of Census 2000: Design and Methodology."
- Whitford, D. (2008). "2010 Census Coverage Measurement: Block Cluster Delineation." DSSD 2010 Census Coverage Measurement Memorandum Series #2010-C-03.

Attachment 1: High-Level Workflow of Key CCM Operations



*Shaded areas indicate field operations

Attachment 2: Codes Used in 2010 CCM Housing Unit Matching

Draft
9/29/2008

Computer Matching

M=Match

P=Possible Match

Pre-processing

M = The CCM and census addresses match.

P = The CCM and census addresses are possible matches.

NI = The CCM address is not matched to a census address.

NE = The census housing unit is not matched to a CCM address.

KE = The census housing unit does not have enough information to send to followup. All address fields and MSN, with the exception of city, state, and zip, are blank.

Q2 = The census address within the cluster is a GQ or Other Living Quarters (OLQs).

N3 = The census housing unit within the surrounding blocks is not matched to a CCM address.

Q3 = The census address is a GQ or an OLQ within the surrounding blocks and is not matched to a CCM address.

DI = The CCM address is a possible duplicate of another CCM address.

DE = The census address is a possible duplicate of another census address.

BFU Clerical Matching

Match Codes

M = The CCM and census addresses are matched.

P = The CCM and census addresses are possible matches. There is not enough information to assign a match with confidence.

NI = The CCM address is not matched to a census address.

NE = The census housing unit is not matched to a CCM address.

- Q2 = The census address within the cluster is a GQ or OLQ.
 N3 = The census address is a housing unit within the surrounding blocks and is not matched to a CCM address.
- Q3 = The census address is a GQ or an OLQ within the surrounding blocks and is not matched to a CCM address.
- KE = The census address does not have enough information to send to followup. All address fields and MSN, with the exception of city, state, and zip, are blank or too ambiguous (e.g., "white house" when there are 6 other "white houses" listed in the same area).
- DI = The CCM address is a possible duplicate with another CCM address. A followup interview is required to determine if the CCM address is actually a duplicate of another CCM address.
- DE = The census address is a possible duplicate with another census address. A followup interview is required to determine if the census address is actually a duplicate of another census address.
- ZM = The CCM map spot number associated with an address is in error. A delete code of ZM is entered for that CCM address. This code removes the address from further CCM processing.
- Y1, Y2= Spare codes for linked cases.
 Y3, Y4= Spare codes for unlinked CCM cases
 Y5, Y6= Spare codes for unlinked census cases

Review Note – A technician can put a review note on a case that sends it to an analyst. The note will be resolved before analyst closeout. CCM GQ Flag

Note, these flags are set only on units that have a GQ flag from keying=1 (yes) indicating that the unit is a possible GQ (GQ Flag=Q) or if the unit matches a census GQ. Otherwise, the flag needs to be blank. Only analysts can set this flag (unless we want to allow technicians to be able to set the flag because we want to work these early)

- ~~Y = This unit is confirmed as a group quarters via an internet search by an analyst. No followup is needed for GQ status.~~
- ~~Q = It is questionable whether this unit is a group quarters or not. A followup interview is needed to determine the status of the group quarters.~~
- ~~N = This unit is not a group quarters as confirmed via an internet search by an analyst. No followup is needed for GQ status.~~

AFU Clerical Matching

Match Codes

- M = The CCM and census addresses match and the units are located in the sample block cluster.
- MS = The CCM and census addresses match and the units are located in the surrounding blocks to the sample block cluster. The CCM address is a geocoding error and the census unit is a correct enumeration.
- ME = The CCM and census addresses match and the units are located outside surrounding blocks to the sample block cluster. Both the CCM address and the census unit are geocoding errors.
- MU = The CCM and census addresses match and there is not enough information on the followup form to confirm this match as a housing unit with certainty. The followup interview was not done, was incomplete, was never sent, had contradictory information, or was a noninterview. The status or location of the unit is unknown.
- MQ = The CCM and census addresses are linked and are a GQ.
- MZ = The CCM and census addresses are linked and do not refer to a housing unit at the time of the followup interview. For example, the housing unit burned or the mobile home moved. This address stays on the PEL.¹
- MX = The CCM and census addresses do not refer to a unit that should have been listed. For example, the units merged, the address is commercial property at the time of the followup interview, or the map spot was in error. This address is not on the PEL.
- CI = The CCM housing unit existed as a housing unit at the time of the followup interview and is correctly geocoded in the block cluster. The housing unit is not found in the census.
- CE = The census housing unit existed as a housing unit at the time of the followup interview and is correctly geocoded in the block cluster. The housing unit is not found in the CCM.
- ZI = The CCM address did not refer to a housing unit at the time of the followup interview. For example, the housing unit burned or the mobile home moved. This address stays on the PEL.

¹ The Preliminary Enhanced List is the list of CCM and census units eligible to be selected to receive the CCM Person Interview.

- XI = The CCM address did not refer to a unit that should have been listed. For example, the units merged, the address is commercial property at the time of the followup interview, or the map spot was in error. This address is not on the PEL.
- ZQ = The CCM address is a group quarters.
- EE = The census housing unit is erroneously listed on the UCM, because the address is not a housing unit in the sample block cluster or the surrounding blocks at the time of the followup interview. For example, the housing unit burned or the mobile home moved, the address is commercial property, or the address is nonexistent within the sample cluster.
- GQ = The census housing unit is erroneously listed on the UCM, because the address is a GQ at the time of the followup interview.
- GI = The CCM address existed as a housing unit at the time of the followup interview, but is incorrectly listed in the block cluster. The address is a CCM geocoding error.
- GS = The census housing unit existed as a housing unit at the time of the followup interview, but is located in the surrounding blocks to the block cluster. This housing unit is correctly enumerated for this block cluster.
- GE = The census housing unit existed as a housing unit at the time of the followup interview, but is located beyond the surrounding blocks to this block cluster. This housing unit is erroneously enumerated in this block cluster, because of a geocoding error.
- DE = The census housing unit is erroneously enumerated in the census. The reason for erroneous enumeration is the address is duplicated in the census.
- DI = The housing unit should not have been listed in the CCM. This address is a duplicate of another CCM address. This address is removed from further processing for the CCM.
- DA = The CCM housing unit is a duplicate identified in AFU that has not been field-confirmed by IHUFU. It will be sent to FHUFU for a field interview. It will remain on the PEL and is eligible for a PI. These are either in a cluster that skipped BFU or was not found in BFU matching.
- DF = The census housing unit is a duplicate identified in AFU that has not been field-confirmed by IHUFU. It will be sent to FHUFU for a field interview if it is in the E sample. These are either in a cluster that skipped BFU or was not found in BFU matching.

- UI = Not enough information on the followup form to assign a code to the nonmatched CCM housing unit with certainty. The followup interview was not done, was incomplete, was never sent, had contradictory information, or was a noninterview.
- UE = Not enough information on the followup form to assign a code to the census nonmatched housing unit with certainty. The followup interview was not done, was incomplete, was never sent, had contradictory information, or was a noninterview.
- Q2 = The census address within the cluster is a GQ or OLQ.
- N3 = The census address is a housing unit within the surrounding blocks is not matched to a CCM address.
- Q3 = The census address is a GQ or an OLQ within the surrounding blocks and is not matched to a CCM address.
- Y1, Y2= Spare codes for linked cases.
- Y3, Y4= Spare codes for unlinked CCM cases
- Y5, Y6= Spare codes for unlinked census cases

Review Note – A technician can put a review note on a case that sends it to an analyst. The note will be resolved before analyst closeout. ~~Any CCM unit with a CCM GQ Flag=Q or Y.~~

Relisting Calculation

$$\text{Relisting \%} = \frac{(MS + GI + ME) * 100}{(M + CI + MS + GI + ME)}$$

If Relisting % is greater than 80, then the cluster should be sent to relisting.

Note—This is the same calculation as 2000, with the addition of MS and ME (new codes)
Attachment 3: Codes Used in 2010 Person Matching

Status Codes

Residence Status Codes				
Residence Status Code	Meaning	Definition	Assigned by Computer	Assigned Clerically
<i>P-sample</i>				
R	Review	This code is assigned by PI Post Processing to cases that need clerical review.	X	
N	Nonmover	People who are Census Day and Interview Day residents of the same P-sample housing unit.	X	X
I	Inmover	People who are Interview Day residents of the P-sample housing unit but are Census Day residents of a different in-scope housing unit. (An in-scope housing unit is one that could have been included in the P sample.)		X
O	P-sample outmover	People who are Census Day residents of the P-sample housing unit but are not Interview Day residents of the same P-sample housing unit. They moved out after Census Day to a unit that was not eligible to be in the P sample. These people include outmovers to a GQ, a housing unit outside the U.S., and people now experiencing homelessness. This also includes people who have died between Census Day and Interview Day.	X	X
U	Unclassified	People whose residence cannot be classified because not enough information was obtained in the PI interview (BFU stage) or the PFU interview (AFU stage) to determine whether they are residents of the P-sample housing unit on Interview Day. For AFU, when certain questions in Section C of the PFU form, about other places where a person could have been counted, are answered “Don’t Know” or “Refused”, this results in an unclassified	X	X

		residence status code.		
<i>Non-P-sample</i>				
A	Never Resident	People who may or may not be Census Day residents of the P-sample housing unit but who have another (or alternate) residence where they should be counted according to the Census Residence Rule.		X
Residence Status Codes				
Residence Status Code	Meaning	Definition	Assigned By Computer	Assigned Clerically
Z	Non-P-sample outmover	People who are Census Day residents of the P-sample housing unit but are not Interview Day residents of the same P-sample housing unit. They moved out after Census Day to a housing unit that was eligible to be in the P sample.		X
X	Out of scope	People who are Census Day residents of a group quarter or a housing unit that is outside the test sites. This includes people who were experiencing homelessness on Census Day, were born after Census Day, or died before Census Day.	X	X
C	Unclassified outmover	People who are outmovers but we are unsure if they moved to a GQ (e.g., they are either a P-sample outmover or a non-P-sample outmover).		X

Computer Match Codes – from PSDB

M – The PI record and the census record match.

P – The PI record and census record possibly match.

Pre-Clerical Codes – assigned by PerMaRCS

Linked

M – The PI record and the census record match.

P – The PI record and census record possibly match.

Unlinked PI

NP – The P-sample person record with status code=nonmover, P-sample outmover, or unclassified is unlinked.

NN – The non-P-sample person record (status code=out-of-scope, never resident, non-P-sample

outmover, or unclassified outmover) is unlinked.

NI – The inmover person record (status code=inmover) is unlinked.

Unlinked Census

NE – The E-sample enumeration is unlinked.

N2 – The person record is in a HU and not part of the E sample, is located within the sample block, but is not linked. This code is applied to those people subsampled out of the E sample in the sample block.

Q2 – The person record is in a GQ and not part of the E sample, is located within the sample block, but is not linked. This code is applied to GQ enumerations in the sample block.

N3 – The HU person is located in the surrounding blocks.

Q3 – The GQ person is located in the surrounding blocks.

N4 – The HU person is located beyond the surrounding blocks.

Q4 – The GQ person is located beyond the surrounding blocks.

Duplicate PI

DM – The PI record is a strong (matched) duplicate of another PI record. (Note – these are only within-sample-block cluster duplicates because we do not search for between sample block cluster duplicates in the PI).

DP – The PI record is a possible (weak) duplicate of another PI record. (Note – these are only within-sample-block cluster duplicates because we do not search for between sample block cluster duplicates in the PI).

Duplicate Census

DS – The census record is a strong (matched) duplicate of another census record.

DW – The census record is a possible (weak) duplicate of another census record.

Sufficiency Status Code (assigned to all PI and all census people)

S = Person record has a complete, valid name for followup (2 characters in the first/middle name fields, 2 characters in the last name field)

I = Person record does not have a complete, valid name for followup

Clerical BFU Codes

Linked (assigned to all linked PI people)

M – The PI record and the census record match.

P – The PI record and census record possibly match.

Unlinked PI

NP – The P-sample person record with status code=nonmover, P-sample outmover, or unclassified is unlinked.

NN – The non-P-sample person record (status code=out-of-scope, never resident, non-P-sample outmover, or unclassified outmover) is unlinked.

NI – The inmover person record (status code=inmover) is unlinked.

FP – The P-sample person is fictitious. The record refers to a pet or a deceased person.

Unlinked Census

NE – The E-sample enumeration is unlinked.

FE – The E-sample person is fictitious. The record refers to a pet or a deceased person.

GE- The E-sample person is in a housing unit located outside the surrounding blocks. It does not need followup.

N2 – The person record is in a HU and not part of the E sample, is located within the sample block, but is not linked. This code is applied to those people subsampled out of the E sample in the sample block.

Q2 – The person record is in a GQ and not part of the E sample, is located within the sample block, but is not linked. This code is applied to GQ enumerations in the sample block.

N3 – The HU person is located in the surrounding blocks.

Q3 – The GQ person is located in the surrounding blocks.

N4 – The HU person is located beyond the surrounding blocks.

Q4 – The GQ person is located beyond the surrounding blocks.

Duplicate PI

DM – The PI record is a strong (matched) duplicate of another PI record. (Note – these are only

within-sample-block cluster duplicates because we do not search for between sample block cluster duplicates in the PI.)

DP – The PI record is a possible (weak) duplicate of another PI record. (Note – these are only within-sample-block cluster duplicates because we do not search for between sample block cluster duplicates in the PI.)

Duplicate Census

DS – The census record is a strong (matched) duplicate of another census record.

DW – The census record is a possible (weak) duplicate of another census record.

Inmover Address Code (IAC or Good/Bad/Ugly Code)

This code is assigned to all inmover addresses to assess the quality of the address information collected in PI.

G – Good. At the geocoded inmover address, the neighbors match, the cohabitants match, or the cross streets match.

B – Bad. At the geocoded inmover address, we have respondent information for neighbors, cohabitants, or cross-streets and they do not match.

U – Ugly. We do not have respondent information on neighbors, co-habitants, or cross-streets or the address was ungeocodable.

Sufficiency Status Code (assigned to all PI and all census people)

S = Person record has a complete, valid name for followup (2 characters in the first/middle name fields, 2 characters in the last name field and is not something like “Mrs. Smith” or “Donald Duck”).

I = Person record does not have a complete, valid name for followup.

AFU Match Codes

Status Code = Nonmover

M – Matched resident of the sample block cluster on Census Day and Interview Day. The P sample and census person are matches. The P-sample person is a resident of the sample block cluster on both Interview Day and Census Day. If linked to an E-sample person, the E-sample person is a correct enumeration and the P-sample person is

considered a matched resident for net error. If linked to a record outside the sample block cluster search area, the P-sample person is a nonmatched resident for net error.

MN – Matched unresolved case due to PFU noninterview. The PI and census record are matches. During the PFU, we are unable to complete an field interview. The case is unresolved.

P – Possibly matched resident of the sample block cluster on Census Day and Interview Day. The P-sample and census person are possible matches. The P-sample person is a resident of the sample block cluster on both Interview Day and Census Day.

PN – Possibly matched unresolved case due to PFU noninterview. The PI and census record are possible matches. During the PFU, we are unable to complete an field interview. The case is unresolved.

N – Nonmatched resident of the sample block cluster on Census Day and Interview Day. The P-sample person is a resident of the sample block cluster on both Census Day and Interview Day and is a nonmatch for net error.

NN – Nonmatched unresolved case due to PFU noninterview. During the PFU, we are unable to complete a field interview. The case is unresolved.

Status Code = P-sample Outmover

M – Matched resident of the sample block cluster on Census Day. The P-sample and census person are matches. The P-sample person is a resident of the sample block cluster on Census Day. If linked to an E-sample person, the E-sample person is a correct enumeration and the P-sample person is considered a match for net error. If linked to a record outside the sample block cluster search area, the P-sample person is a nonmatch for net error.

MN – Matched unresolved case due to PFU noninterview. The PI and census record are matches. During the PFU, we are unable to complete an field interview. The case is unresolved.

P – Possibly matched resident of the sample block cluster on Census Day. The P-sample and census person are possible matches. The P-sample person is a resident of the sample block cluster on Census Day.

PN – Possibly matched unresolved case due to PFU noninterview. The PI and census record are possible matches. During the PFU, we are unable to complete an field interview. The case is unresolved.

N – Nonmatched resident of the sample block cluster on Census Day. The P-sample person is a resident of the sample block cluster on both Census Day and is a nonmatch for net error.

NN – Nonmatched unresolved case due to PFU noninterview. During the PFU, we are unable to complete an field interview. The case is unresolved.

Status Code = Inmover

M - Matched inmover. The P-sample and census person are matches. The P-sample person is an inmover and was a resident of the sample block cluster on Interview Day and a resident of an in-scope¹ housing unit on Census Day. If the census person is in the inmover search area, the P-sample person is considered a match for net error. If the census person is not in the inmover search area, the P-sample person is considered a nonmatch for net error.

MN – Matched unresolved case due to PFU noninterview. The PI and census record are matches. During the PFU, we are unable to complete an field interview. The case is unresolved.

P = Possibly matched inmover. The P-sample and census person are possible matches. The census person is in the inmover search area.

PN – Possibly matched unresolved case due to PFU noninterview. The PI and census record are possible matches. During the PFU, we are unable to complete an field interview. The case is unresolved.

N - Nonmatched inmover. The P-sample person is a resident of the sample block cluster on Interview Day and a resident of an in-scope housing unit on Census Day and is a nonmatch for net error.

NN – Nonmatched unresolved case due to PFU noninterview. During the PFU, we are unable to complete an field interview. The case is unresolved.

Status Code = Never Resident

M– Matched nonresident of the sample block cluster on Census Day and Interview Day. The PI and census person are matches. The PI person should be counted at their alternate address or nationwide address (if more than one address, then we need to collect which address for component error). If linked to an E-sample person, the E-sample person is an erroneous enumeration and should also be counted at the alternate address.

MN – Matched unresolved case due to PFU noninterview. The PI and census record are matches. During the PFU, we are unable to complete an field interview. The case is unresolved.

P – Possibly matched nonresident of the sample block cluster on Census Day and Interview Day. The PI and census person are possible matches.

¹An “in-scope” housing unit is one that could be sampled in the P-sample.

PN – Possibly matched unresolved case due to PFU noninterview. The PI and census record are possible matches. During the PFU, we are unable to complete a field interview. The case is unresolved.

N – Nonmatched nonresident of the sample block cluster on Census Day and on Interview Day. The PI person should be counted at their alternate or nationwide address (if more than one address, then we need to collect which address for component error).

NN – Nonmatched unresolved case due to PFU noninterview. During the PFU, we are unable to complete a field interview. The case is unresolved.

Status Code = Non-P-sample Outmover

M– Matched nonresident of the sample block cluster on Census Day and Interview Day. The PI and census person are matches. The PI person should be counted at their alternate address or nationwide address (if more than one address, then we need to collect which address for component error). If linked to an E-sample person, the E-sample person is an correct enumeration.

MN – Matched unresolved case due to PFU noninterview. The PI and census record are matches. During the PFU, we are unable to complete a field interview. The case is unresolved.

P – Possibly matched resident of the sample block cluster on Census Day and Interview Day. The PI and census person are possible matches.

PN – Possibly matched unresolved case due to PFU noninterview. The PI and census record are possible matches. During the PFU, we are unable to complete a field interview. The case is unresolved.

N – Nonmatched nonresident of the sample block cluster on Census Day and on Interview Day. The PI person should be counted at their alternate or nationwide address (if more than one address, then we need to collect which address for component error).

NN – Nonmatched unresolved case due to PFU noninterview. During the PFU, we are unable to complete a field interview. The case is unresolved.

Status Code= Out of Scope

M – Matched out-of-scope person. The PI and census person are matches. The PI person should not have been counted in the housing unit universe and should have been counted in a GQ, was homeless, was born after Census Day, or died before Census Day. If linked to an E-sample person, the E-sample person is an erroneous enumeration for net error and

for component error. The system must collect which address (if any) where the person should be counted.

MN – Matched unresolved case due to PFU noninterview. The PI and census record are matches. During the PFU, we are unable to complete an field interview. The case is unresolved.

MF – Matched fictitious person. The PI and census person are matches. The PI person was determined to be fictitious after interviewing 3 knowledgeable respondents during the person followup interview. If matched to an E-sample person, the E-sample person is an erroneous enumeration.

P – Possibly matched out-of-scope person. The PI and census person are possible matches.

PN – Possibly matched unresolved case due to PFU noninterview. The PI and census record are possible matches. During the PFU, we are unable to complete an field interview. The case is unresolved.

PF – Possibly matched fictitious person. The PI and census person are possible matches. The PI person was determined to be fictitious after interviewing 3 knowledgeable respondents during the person followup interview. If linked to an E-sample person, the E-sample person is an unresolved enumeration.

N – Nonmatched out-of-scope person. The PI person should not have been counted in the housing unit universe and should have been counted in a GQ, was homeless, was born after Census Day, or died before Census Day. The system must collect which address (if any) where the person should be counted.

NN – Nonmatched unresolved case due to PFU noninterview. During the PFU, we are unable to complete an field interview. The case is unresolved.

GP – The P-sample case is a geocoding error and is rostered in an address located outside the sample block cluster search area but was listed inside the sample block cluster search area.

FP – P-sample fictitious person. The PI person was determined to be fictitious after interviewing 3 knowledgeable respondents during the person followup interview.

Duplicate Codes (any status code)

DM – The PI record is a strong (matched) duplicate of another PI record. (Note – these are only within-sample-block duplicates because we do not search for between-sample-block duplicates in the PI).

DP – The PI record is a possible (weak) duplicate of another PI record. (Note – these are only

within-sample-block duplicates because we do not search for between-sample-block duplicates in the PI).

Status Code = Unclassified

MU – The matched PI record is unresolved. The record was unresolved during BFU and remains unresolved after PFU or the record was resolved during BFU, an interview was obtained, and not enough information was gathered to classify the person with a resolved status code.

P – The possibly matched PI record is unresolved. The record was unresolved during BFU and remains unresolved after PFU or the record was resolved during before followup, an interview was obtained, and not enough information was gathered to classify the person with a resolved status code.

NU – The nonmatched PI record is unresolved. The record was unresolved during BFU and remains unresolved after PFU or the record was resolved during BFU, an interview was obtained, and not enough information was gathered to classify the person with a resolved status code.

Status Code = Unclassified Outmover

MU – The matched PI record was a Census Day resident of the sample block cluster and moved out. It is not known if the person moved to an in-scope HU or elsewhere (i.e., we cannot tell if the person is a P-sample outmover or a non-P-sample outmover).

P – The possibly matched PI record is unresolved and the PI person is either a non-P-sample outmover or a P-sample outmover.

NU – The nonmatched PI record is unresolved and the PI person is either a non-P-sample outmover or a P-sample outmover.

Unlinked to PI people Census Cases

CE – The census person is a correct enumeration. This code can be applied to cases outside the sample block cluster when an E-sample person is a duplicate of a case outside the sample block cluster search area and the correct enumeration is outside the cluster. The E-sample person would then be a duplicate.

UE – The census person is an unresolved enumeration. Either the PFU interview was a noninterview or the respondent did not know the answers to the PFU questions in order to resolve the person. This code can be applied to cases outside the sample block cluster when an E-sample person is a duplicate of a case outside the sample block cluster search area and the unresolved enumeration is outside the sample block cluster. The E-sample person would then be a duplicate.

- EA – The unlinked census person should be counted outside of this sample block cluster in an in-scope housing unit. Examples of these type of people include those who move into the sample block cluster after Census Day from an in-scope housing unit, those who move out of the sample block cluster before Census Day into an in-scope housing unit, a person who cycles between two or more in-scope housing units and this sample block cluster is not their residence according to Census Residence Rules. (This is the analog to a P-sample never resident person or in-mover.)
- EX – The unlinked census person should be counted in a GQ on Census Day, was homeless on Census Day, died before Census Day, was born after Census Day. (This is the analog to a P-sample out-of-scope person.)
- FE – E-sample fictitious person. The census person was determined to be fictitious after interviewing 3 knowledgeable respondents during the person followup interview.
- GE – The E-sample case is a geocoding error and is rostered in an address located outside the sample block cluster search area but was listed inside the sample block cluster.
- DS – The census record is a strong (matched) duplicate of another census record. This code can be used for within-sample-block cluster search area duplicates or for duplicates sent to PFU that were not nation-site duplicates. If a duplicate found during the nationwide computer matching was mistakenly sent to PFU because an alternate address was not geocoded properly in BFU, then this code can also be used.
- DW – The census record is a possible (weak) duplicate of another census record. This code can be used for within-sample-block cluster search area duplicates or for duplicates sent to PFU that were not duplicates. If a duplicate was mistakenly sent to PFU because an alternate address was not geocoded properly in BFU, then this code can also be used.
- DC – Field-confirmed nationwide duplicate. If a nationwide duplicate is confirmed by either the within-sample cluster block search area case or the nationwide phone case, then this code should be applied to the record that is erroneously counted.
- DN – Unconfirmed nationwide duplicate with evidence– This code is applied when a duplicate is not field-confirmed but the clerical matcher has evidence that the record is a true duplicate.
- DU – Unconfirmed nationwide duplicate without evidence. This code is applied when a nationwide duplicate is not field-confirmed by either PFU case and there is no evidence of a true duplicate. The DU code is retained even though the clerical matcher does not consider the record a duplicate.

Inmover Address Code (IAC or Good/Bad/Ugly Code)

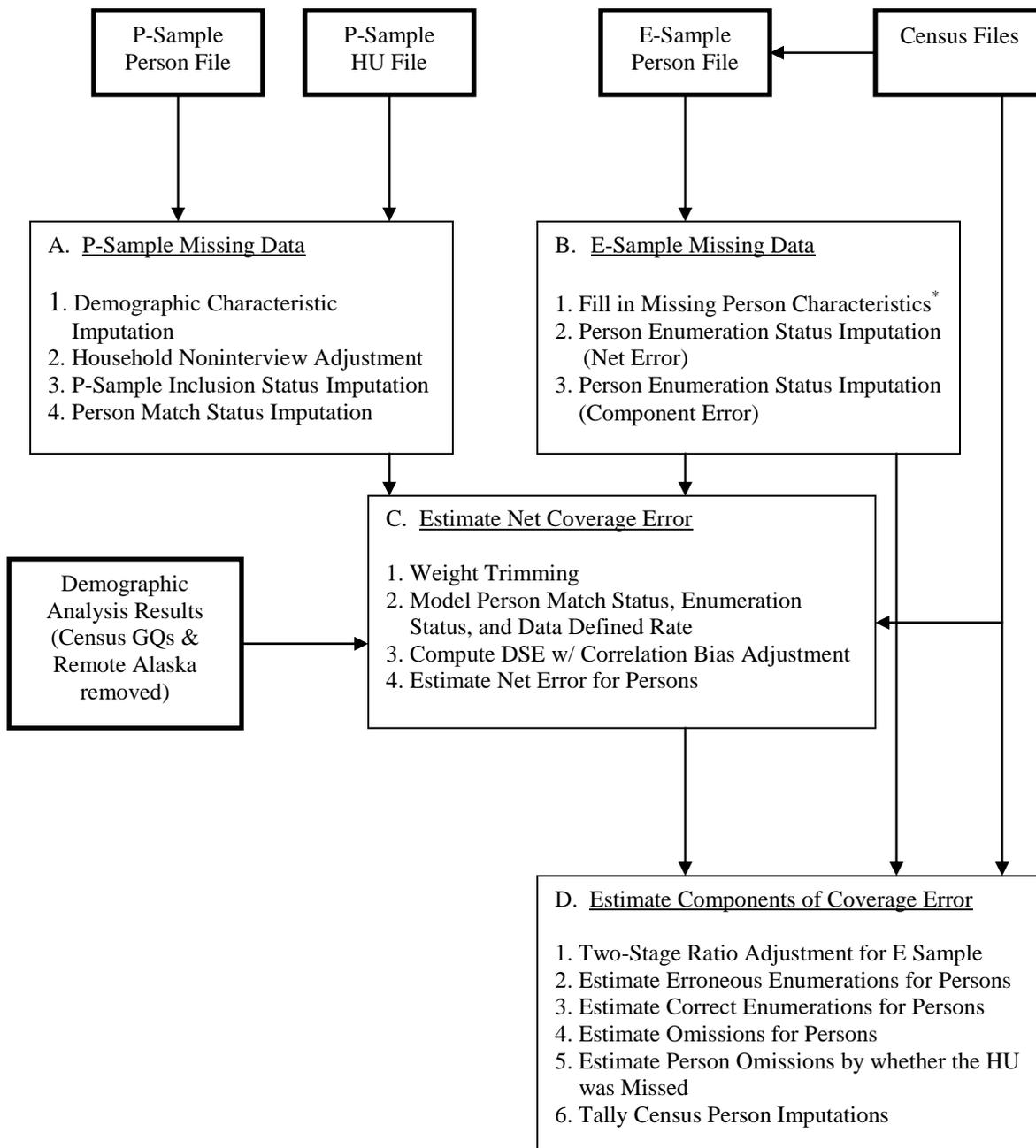
This code is assigned to all in-mover addresses to assess the quality of the address information collected in PI/PFU.

- G – Good. At the geocoded in-mover address, the neighbors match, the cohabitants match, or the cross streets match.
- B – Bad. At the geocoded in-mover address, we have respondent information for neighbors, cohabitants, or cross-streets and they do not match.
- U – Ugly. We do not have respondent information on neighbors, cohabitants, or cross-streets or the address was un-geocodable.

Attachment 4: Search Areas and CCM Person Matching Operations

Search area	Residence status of people associated with this search area	What happens with respect to this search area during computer matching?	What happens with respect to this search area during clerical matching?
Sample Block & Surrounding Blocks	Nonmovers, P-sample Outmovers	Person records are linked across housing units. Links are then modeled. Links selected to be matches or possible matches within the sample block search area are considered sample block links.	Analysts search possible matches and nonmatches within the sample block search area for additional links.
Inmover Address & Surrounding Blocks	Inmovers	Person records are linked across housing units. Links are then modeled. Links selected to be matches or possible matches within the inmover address search area are considered inmover address links.	Analysts search possible matches and nonmatches within the inmover block search area for additional links.
Alternate Address & Surrounding Blocks	Nonmovers, Inmovers, P-sample Outmovers, Non-P-sample outmovers	Person records are linked across housing units. Links are then modeled. Links selected to be matches or possible matches within the alternate address search area are considered alternate address links.	Analysts search possible matches and nonmatches within the alternate address block search area for additional links.
Nationwide Address	Nonmovers, Inmovers, P-sample Outmovers, Non-P-sample outmovers	Person records linked across housing units which are neither associated with an inmover address nor an alternate address are classified as nationwide links.	Analysts search for other possible matches and duplicates within the MAF Identification Number (MAF ID) of the nationwide link.

Attachment 5: Overview of CCM Estimation – Persons in Housing Units



* Missing person characteristics are equivalent to the values assigned via the Census Edit and Imputation Process

Attachment 7: CCM Glossary

Census Day residence status	A determination of the residence status of the PI person at the sample address on Census Day. The result may be: a Census Day resident, a Census Day nonresident, or an unclassified Census Day residence status.
components of coverage error	The four components of census coverage error are: omissions, erroneous enumerations, correct enumerations, and whole-person census imputations. Whole-person census imputations are applicable only for person estimates.
correct enumeration	Refers to a census enumeration in the correct universe that was counted once and only once.
E sample	A sample of census enumerations. The E sample is comprised of housing units and census person enumerations in housing units in the same sample block clusters as the P sample. After matching and followup, the E sample yields information to estimate correct and erroneous enumerations, as well as the correct enumeration rate.
erroneous enumeration	Refers to a census enumeration that is a duplicate of another census enumeration or an enumeration that should not have been included in the universe.
group quarters (GQ)	A living quarters in which people live or stay that is normally owned or managed by an entity or organization providing services for the residents. GQs include residential treatment centers, hospices, hospital wards, college residence halls, nursing homes, prison wards, juvenile institutions, dormitories, workers' dormitories, and facilities for people experiencing homelessness.
housing unit (HU)	A living quarters in which the occupants live separately from any other individuals in the building and have direct access to their living quarters from outside the building or through a common hall.
inmover	A person who is an Interview Day resident at the sample address and a Census Day resident at a different address.
Interview Day	The day the coverage measurement PI is conducted at a particular sample address.
net coverage error	Reflects the difference between the estimated population (or subpopulation) total obtained through dual system estimation and the census count for that population (or subpopulation). A positive net error indicates a census undercount, while a negative net error indicates a census overcount.
nonmover	A person who is an Interview Day resident and a Census Day resident of a sample address.
omission	A person or housing unit that should have been a census enumeration in the universe, but was not. Census Residence Rules are used to determine where persons should be enumerated in the census.
outmover	A person who is not an Interview Day resident at a particular sample address, but was a Census Day resident at that sample address. A P-sample outmover is an outmover who <u>should</u> be included in the P sample because, on Interview Day, he/she resides in a residence not eligible to be in the P sample. A non-P-sample outmover is an outmover who <u>should not</u> be included in the P sample, because on Interview Day he/she resides in another residence eligible to be in the P sample.
P sample	A sample of housing units and persons in housing units obtained independently from the census. After matching and followup, the P sample provides information about the housing units or people who were missed in the census.