October 1, 2012

2010 CENSUS PLANNING MEMORANDA SERIES

No. 241

MEMORANDUM FOR   The Distribution List

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

Subject: 2010 Census Evaluation to Assess Effect of Census Coverage Measurement (CCM) Search Area and Census Address List Formation Rules on CCM Estimates Report

Attached is the 2010 Census Evaluation to Assess Effect of Census Coverage Measurement (CCM) Search Area and Census Address List Formation Rules on CCM Estimates Report. The Quality Process for the 2010 Census Test Evaluations, Experiments, and Assessments was applied to the methodology development and review process. The report is sound and appropriate for completeness and accuracy.

If you have any questions about this document, please contact Rachel Bray at (301) 763-2631.

Attachment

2010 Census Evaluation to Assess the Effect of Census Coverage Measurement Search Area and Census Address List Formation Rules on Census Coverage Measurement Estimates Report

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

Rachel Bray
Decennial Statistical Studies Division

(This Page Intentionally Left Blank)

Table of Contents

List of Tables

Executive Summary

The 2010 Census Coverage Measurement Program (CCM) conducted an independent listing of housing units in sampled block clusters called the P sample (population sample) of housing units. The P sample was then matched to the 2010 Census address list in the sampled blocks clusters, and the first ring of blocks surrounding the block clusters. The 2010 CCM classified 6,223 of 171,217 of the P-sample housing units in the United States as nonmatches to the 2010 Census address list in the sample block cluster and first ring of surrounding blocks. Of these housing units, 2,030 were classified nonmatches by CCM Estimation because they matched to a housing unit that was deleted from the final 2010 Census address list. The remaining 4,193 housing units were classified as nonmatches because they were missing from the 2010 Census address list.

As a part of the Evaluation of Address Frame and Address Quality, the 4,193 missing housing units were searched for on an expanded address list. The expanded address list included addresses within five kilometers of the sample block cluster, in addition to addresses on the Census Bureau's Master Address File that were not in the 2010 Census such as addresses listed as businesses.

The number of missing housing units that were matched to the expanded address list is small. Of the 4,193 housing units that CCM classified as missing from the 2010 Census address list, only 150 were found during the matching to the expanded address list, 15 of which were matched to units deleted from the final 2010 Census address list ("census deletes"). Three housing units remained unresolved. The remaining 4,040 housing units persisted as nonmatches, even after expanding the address list geographically and including units not on the 2010 Census address list.

This evaluation looks at the 135 missing housing units that were found in the Evaluation of Address Frame and Address Quality matching to the expanded address list, that were not matched to census deletes. This evaluation reports the level of geography, relative to the CCM search area of the sample block cluster and first ring of surrounding blocks, at which the housing units were found. This evaluation also reports whether the missing addresses were matched to the 2010 Census address list, or to the additional addresses on the Master Address File. Finally, under some basic assumptions about the relationship between person and housing unit matching, this evaluation reports the potential impact of these two types of errors in the address list used in matching on CCM household population coverage estimates. This evaluation reports this information by addressing the following research questions.

1. The CCM search area consists of the sample block cluster plus the first ring of surrounding blocks adjacent to the block cluster. How many housing units on the final 2010 Census address list were geocoded in error to a block outside of the CCM search area?

Of the 135 missing housing units matched to the extended address list, 100 were found on the 2010 Census address list. Of these 100 housing units, 24 were located within the CCM search area, while 76 were located outside of the CCM search area but within three kilometers of the sample block cluster. These 76 housing units were geocoded in error, by the 2010 Census, to a

block outside of the CCM search area. There were no units found beyond three kilometers outside of the CCM search area.

2. How many CCM P-sample housing units were correctly coded as missing from the 2010 Census address list, but found on the Census Bureau's Master Address File?

Of the 135 missing housing units matched to the extended address list, 35 were found on the Master Address File but not the 2010 Census address list. Of these 35 units, 34 were found in the CCM search area, while only one was found outside of the search area but within one kilometer of the sample block cluster. No housing units were found beyond one kilometer but within five kilometers of the sample block cluster.

3. What is the potential impact of errors in the address list on the CCM estimates of coverage for the household population?

The potential impact of these errors in the address list on CCM estimates of coverage for the household population is low. In order to answer this question, several alternative dual system estimates for persons were computed, under the hypothetical situations where the errors in research questions (1) and (2) were corrected, under the assumption that the persons within a household would have the same outcome as the housing unit. The alternative dual system estimates showed a maximum decrease of about 8,000 persons or 0.003 percent, from the baseline dual system estimate. All of the alternative dual system estimates were less than the baseline estimate.

After the alternative dual system estimates were computed, they were used in calculations of the percent net undercount to determine how much net coverage error estimation would have been impacted by correcting for errors in the address list. The alternative percent net error estimates use the alternative dual system estimates along with a 2010 Census count that was increased in the same way as the counts and estimates used in the dual system estimate calculations. The largest increase in the percent net overcount was 0.02 percent. This measure was very robust to small changes in the dual system estimate and census count, and thus to errors in the address list used in matching.

One of the components of census coverage, the estimate of omissions, also uses the dual system estimate of the population to determine how many people were missed in the 2010 Census. The alternative omissions estimates were calculated using each alternative dual system estimate and the correct enumeration count corresponding to that estimate. The largest change in the estimate of omissions was a decrease of about 150,000 people.

Based on the evidence that refining the address list to include additional units on the Master Address File or expanding the search area geographically would result in little change to CCM estimates of coverage, we recommend no changes to the current search area or address list formation rules for CCM estimation purposes.

1. Introduction

The purpose of this study is to assess the impact of errors on the 2010 Census address list on the 2010 Census Coverage Measurement (CCM) estimates of coverage error. CCM conducts an independent listing of housing units in sample areas, which is referred to as the population sample, or P sample. The CCM then matches the P sample to a list of census addresses within the sample block cluster and surrounding ring of blocks. The Evaluation of Address Frame Accuracy and Quality (AFAQ) matched P-sample housing units that CCM coded missing from the 2010 Census address list to an expanded address list. The expanded list included addresses that did not make it into the 2010 Census, but are on the Master Address File (MAF); and a geographic area of up to five kilometers outside of the sample block cluster. This evaluation makes use of the AFAQ matching to an expanded address list to provide estimates of errors resulting from the creation of the address list for CCM matching, as well as the source of these errors. This evaluation also explores the potential impact of these errors on the CCM estimates of coverage for the household population.

2. Background

The 2010 CCM program and Census 2000 Accuracy and Coverage Evaluation (A.C.E.) Survey used dual system estimation to estimate the "true" population total and net coverage error for the population and also for housing units. One requirement of the dual system estimation is an independent listing of housing units in sample areas, which in the CCM and A.C.E. is referred to as the P sample. The P-sample housing units are then matched to the 2010 Census address list in sample block clusters and adjacent blocks to determine the match rate, which is used in the computation of the dual system estimate (DSE) of the "true" population total. For more details on the CCM Estimation Methodology, see the 2010 Census Coverage Measurement Estimation Methodology (Mule, 2008). For an overview of the A.C.E. estimation methodology, see the Accuracy and Coverage Evaluation of Census 2000: Design and Methodology (Census, 2004).

The 2010 CCM program also provides a direct estimate, from the Enumeration (E) sample, of the components of census coverage. One component of census coverage for housing units is geocoding error. A housing unit considered a geocoding error in the 2010 Census if it is found outside of the block cluster and adjacent blocks. E-sample housing units that are geocoded to the sample block cluster, but are actually located beyond the surrounding blocks to the sample block cluster are classified as census geocoding errors and contribute to this CCM estimate. However, this estimate of geocoding error may be an underestimate due to limited searching outside of the search area. For the 2010 Census, CCM estimated 0.1 percent of all census housing units to be census geocoding errors, with a standard error of 0.03 percent (Mule and Konicki, 2012).

The Assessment of Addresses on the Master Address File "Missing" in the Census or Geocoded to the Wrong Collection Block (Ruhnke, 2003) performed an extended search for housing units in the A.C.E. that were coded as "missing" from the Census 2000 address list. The A.C.E. coded a housing unit in the P sample as "missing" when it could not be found in Census 2000 in the sample block cluster or blocks adjacent to the block cluster. The evaluation estimated a 4.8 percent geocoding error in Census 2000 to the incorrect block, with a standard error of 0.3 percent, by combining the extended search results and the A.C.E. results. This evaluation did not

make any distinction between units that were geocoded to an adjacent block, and those that were geocoded farther away. The 2010 Census Evaluation of Address Frame Accuracy and Quality (Johnson and Kephart) will provide more in-depth analysis of geocoding error, including estimates of error outside of the block but within the search area.

In addition, the Assessment of Addresses on the Master Address File "Missing" in the Census or Geocoded to the Wrong Collection Block (Ruhnke, 2003) found that some Census 2000 P-sample "missing"-coded housing units were found to have been accounted for on the Census Bureau's MAF but never sent to the decennial version of the MAF (DMAF), for the start of address list preparations. The DMAF was subject to continual refinement throughout census operations. An estimated 54 percent (4,800 of 8,900) of the P-sample housing units coded as "missing" in Census 2000 were eventually accounted for on the MAF. Of those housing units accounted for, 65 percent (3,100 of 4,800) were located in a different block while the remainder were found within the same block to which the A.C.E. had the units geocoded.

The 2010 CCM classified two situations resulting from CCM final housing unit matching as nonmatches. The first, and focus of this evaluation, consists of P-sample housing units that were not found on the 2010 Census in the CCM search area. These housing units were thus missing from the 2010 Census list in the sample block cluster or first ring of surrounding blocks. The second reason that CCM called a housing unit a nonmatch was because the housing unit matched to a record that was on the 2010 Census address list but was deleted during census operations (referred to as "census deletes"). During processing to support CCM estimation, the matches to census deletes were "broken" and these records were coded as nonmatches. Results from the analysis of census deletes are forthcoming (Johnson and Kephart).

The 2010 Census used enhanced methods in constructing its universe file (analogous to the Census 2000 DMAF) – the Universe Control and Management File. The initial Universe Control and Management File was then updated throughout census operations to produce the final 2010 Census address list. The AFAQ evaluation investigates the quality of the 2010 Census address list at several points in time, including after all 2010 Census operations have been completed (Johnson, 2010). The AFAQ evaluation looked for errors in the 2010 Census address list in two ways that are particularly relevant to this study. The first was by performing a search for P-sample housing units in a geographic area beyond the block cluster and surrounding ring of blocks. The AFAQ evaluation searched for these addresses in blocks within one, three, and five kilometer rings concentric to the sample block cluster. The second way in which the AFAQ evaluation looked for errors in the 2010 Census address list used for matching was by searching for the P-sample housing units not found on the 2010 Census address list on the MAF.

This study reports the number of records that the AFAQ evaluation successfully matched to an extended address list along with their geography and whether the units were found in the 2010 Census or just on the MAF. In addition, this evaluation explores, under some simplifying assumptions, the way in which matching these additional addresses would impact the CCM estimates of coverage for the household population.

3. Methodology

3.1 Dual System Estimation and the Match Rate

When a population total is unknown, such as the true population in a nation, one can use a process called dual system estimation to get an estimate of the "true" total. Dual system estimation gets its name from the two systems: the E sample, which is a sample of census enumerations, and the P sample, which is a listing of housing units and people formed independently from the census list. The 2010 CCM and 2000 A.C.E. both used dual system estimation to get estimates of the total household population that should have been counted in the census, and thus of the net error of the census count. The basic formula for the DSE of a population total is given below.

$$DSE_{onecell} = DD \times \frac{CE/E}{M/P} = DD \times \frac{CE}{E} \times \frac{P}{M} \qquad (Equation\ 1)$$

Where:

DD= data-defined total of census people[1]
CE = weighted estimate of correctly enumerated people
E = weighted estimate of E-sample people
P = weighted estimate of P-sample people
M = weighted estimate of person matches

A predefined set of post-strata are used to determine the national DSE. For Census 2000 this was done explicitly through post-stratification, while for the 2010 Census this was accomplished through logistic regression (Mule, 2008). For the purposes of this study, a baseline one-cell national DSE is calculated for the household population from Equation 1. This one-cell DSE is calculated in the same way as in Olson and Sands (2012) and thus, it has the same limitations. The one-cell national DSE is a simplified estimate of the household population that does not use any post-stratification or correlation bias adjustment. As such, the baseline DSE differs slightly from the CCM estimate of the household population. This figure serves as a baseline for comparison to how errors in the 2010 Census address list, and thus in housing unit matching, could have changed the DSE for the household population.

$$DSE_{baseline} = 294{,}709{,}980 \times \frac{\frac{274{,}200{,}057}{260{,}577{,}075}}{\frac{298{,}652{,}586}{286{,}121{,}350}} = 297{,}105{,}140 \quad (Equation\ 2)$$

Part of this evaluation addresses how changing the rules for determining the 2010 Census address list used in matching might impact the CCM estimates of coverage for the household population. Any impact on the DSEs in this context would be the result of adding or subtracting weighted person enumerations from the data-defined count and one or more of four weighted

---

[1] A data-defined person enumeration in the census had two reported characteristics, one of which can be the respondent's name.

estimates: the person matches, the P-sample people, the E-sample people, and the correctly enumerated people. A general formula for possible impacts to the DSE estimates is shown in Equation 3.

$$DSE_{alternative} = (DD + \Delta DD) \times \frac{CE + \Delta CE}{E + \Delta E} \times \frac{P + \Delta P}{M + \Delta M} \qquad (Equation\ 3)$$

Where DD, CE, E, P, and M are defined as before and

ΔDD = change to the data-defined total
ΔCE = change to the weighted total of correctly enumerated
  people
ΔE = change to the weighted total of E-sample people
ΔP = change to the weighted total of P-sample people
ΔM = change to the weighted total of person matches

The changes to the DSE are dependent on both how the housing units were matched (in the CCM search area, or not in the CCM search area) and to which source (2010 Census address list or MAF) the units were matched. The general formula is applicable in all cases, though in specific situations several of the delta terms may be zero, representing no change to the weighted estimate.

3.2 Percent Net Undercount and Omissions

The DSE is used in the computation of two coverage measures associated with the household population. One of these, the percent net undercount, in addition to the DSE, also uses the complete 2010 Census count (Census). This count includes all of the household population that is not in remote Alaska, which corresponds to the in-scope population for CCM. The 2010 Census count includes the data-defined persons as well as whole person imputations (persons with all their characteristics imputed). The second coverage measure is the omissions component of census coverage. This estimate uses the DSE along with the E-sample estimate of correct enumerations (CEs).

The net undercount is the difference between the DSE and the total 2010 Census count of the household population outside of remote Alaska. The percent net undercount (pct_under) is calculated by dividing the net undercount by the DSE. The formula and baseline estimate of percent net undercount are given by Equation 4. Please note that this percent net undercount does not agree with the published 2010 CCM estimate, as it is a simplified estimate which is only used as a point of comparison. For published 2010 CCM estimates please see the 2010 CCM Estimation Report: Summary of Estimates of Coverage for Persons (Mule, 2012).

$$pct\_under_{baseline} = \frac{DSE - Census}{DSE} \times 100 = \frac{297,105,140 - 300,703,438}{297,105,140} \times 100 = -1.21\% \ (Equation\ 4)$$

In addition to estimates of net coverage, one component of census coverage, the estimate of omissions, also uses the DSE. The omissions are calculated by taking the DSE minus the

number of CEs. The formulas and baseline estimate of omissions (omiss) are given by Equation 5.

$$omiss_{baseline} = DSE - CE = \ 297{,}105{,}140 - 274{,}200{,}057 = \ 22{,}905{,}083 \quad (Equation\ 5)$$

For each alternative DSE calculated, a corresponding percent net error and estimate of omissions are also calculated. If the 2010 Census count or count of CEs changed via ΔCE or ΔDD, as defined in Section 3.1, then the change will also be applied to the alternative percent net undercount and omissions estimates.

3.3 Data Overview

In order to answer the questions outlined in the study plan (Mulry et. al, 2011), data were required from five sources: the CCM U.S. P-sample housing unit file, the AFAQ Results file, the AFAQ Geocoding file, the 2010 Census Unedited File (CUF), and the CCM U.S. P-sample person file. The analysis does not include data or results from Puerto Rico.

This evaluation primarily focuses on the P-sample housing units that CCM coded as missing from the 2010 Census address list. CCM matched housing units within a search area of the sample block and its surrounding blocks. The match code that reflects CCM Estimation's post-processing of matching data on the P-sample housing unit file is used to identify the nonmatching housing units (Seiss, 2011). For the purposes of this study, the housing units were first classified as matches or nonmatches. The nonmatches are further classified into those that are missing from the 2010 Census address list, and those that matched to deletes. Table 1 shows the classification of P-sample housing units into the three outcomes.

Table 1: Classification of P-sample Housing Units

P-sample outcome classification	Count
Nonmatches - Missing from 2010 Census address list	4,193
Nonmatches - Broken Matches to deletes	2,030
Remainder of P-sample records not classified as Nonmatches	164,994

Data Source: P-Sample Housing Unit Data and the 2010 CUF

The 4,193 P-sample housing units that are missing from the 2010 Census address list are of interest for this evaluation. For this subset of housing units, the AFAQ evaluation performed additional matching and fieldwork to resolve whether the units were found on the expanded address list. This additional matching and fieldwork are reflected on the two AFAQ evaluation data sources used in this evaluation.

The AFAQ Results file links the newly matched P-sample housing units with the corresponding identifier on the MAF or 2010 Census address list and gives a general match code that simply points to whether the P-sample housing unit was matched, not matched, or remained unresolved. There were only three housing units on the AFAQ Results file that were unresolved.

The AFAQ Geocoding file contains only those records that were matched or potential matches and has a variable that indicates the geography at which the housing unit was matched.

The 2010 CUF Operations Table is used to determine the final status ("delete" or "non-delete") of the census housing unit to which the P-sample housing unit was matched. It is also used to determine whether the housing unit to which the P-sample unit was matched was on the 2010 Census address list. Housing units on the extended address list that are not found on the 2010 CUF are classified as on the MAF and not the 2010 Census address list.

The P-sample person file is merged with the list of newly matched P-sample housing units to obtain the weighted person counts associated with each housing unit. These person weights have undergone weight trimming, per the status imputation and weight trimming specification (Seiss et. al, 2011).

3.4 Questions to be Answered

1) The CCM search area consists of the sample block cluster plus the first ring of surrounding blocks adjacent to the block cluster. How many housing units on the 2010 Census address list were geocoded in error to a block outside of the CCM search area?

2) How many CCM P-sample housing units were correctly coded as "missing" from the 2010 Census address list, but found on the Census Bureau's MAF?

In order to answer the first two questions, the data sources in the preceding section were merged by the housing unit identifiers into a single analysis file to classify the missing housing units into two outcomes. The first outcome gives whether or not the AFAQ evaluation matching resulted in the missing addresses being found on the extended address list. For addresses that were found on the extended address list, the first outcome also gives whether or not the units were matched to census "deletes".

Since CCM classified matches to "deletes" as nonmatches, the missing housing units that matched to census "deletes" have been excluded from further analysis. The second outcome variable is used to then answer question 1 and question 2, using the missing housing units that matched to valid (non-deleted) records on the 2010 Census address list, or on the MAF. This outcome also takes into account the geography at which the missing housing unit was matched.

Although it would have been ideal to use the CCM weights to get nationally representative estimates of these kinds of errors in the 2010 Census address list, not enough errors were found to produce estimates with reliable amounts of precision. The criterion for a reliable amount of precision was a coefficient of variation of less than 20 percent.

3) What impact would fixing the errors in the address list and in matching have on the CCM estimates of coverage for the household population?

In order to determine the impact on the coverage estimates for the household population, the baseline one-cell person DSE in Equation 2 was calculated. This uses data from the P-sample person file, as well as the data-defined person count. Alternative DSEs are determined by adding and subtracting weighted person counts associated with certain housing unit outcomes from the

6

AFAQ evaluation through the ΔDD, ΔM, ΔP, ΔCE, and ΔE terms. The alternative DSEs are then used to calculate a corresponding percent net undercount and omissions estimate. The alternative coverage measures are focused on answering two hypothetical questions concerning errors in the 2010 Census address list.

3a. What if the CCM search area had been extended from the block cluster and surrounding ring of blocks to all blocks within three kilometers of the sample block cluster?

3b. What if the 2010 Census address list had included the additional housing units that the missing-coded housing units matched to on the MAF?

In order to answer these two questions, the analysis file was merged with the P-sample person file to get a count of people (along with their sample weights) associated with the missing, but now matched, housing units.

4. Limitations

- The study plan for this evaluation (Mulry et. al, 2011) outlines matches to the "final UCM" as opposed to the final 2010 Census address list used in final housing unit matching. The final 2010 Census address list is assumed to have addresses that were on the initial UCM, plus addresses that were added through 2010 Census operations, minus addresses that were deleted through 2010 Census operations.
- The study plan also does not take into account the subset of P-sample housing units that matched to housing units deleted from the final 2010 Census address list. In order to not bias the DSE, CCM decided to remove the "deletes" from E-sample processing and to "break" the P-sample matches to "deletes". These housing units are coded as nonmatches, but were typically not followed up during the AFAQ evaluation matching. Thus the housing units that matched to "deletes" are out of scope for this evaluation, and are defined as such in the introduction.
- The AFAQ evaluation address list to which the housing units are matched includes "deletes" from the final 2010 Census address list that may or may not have been on the MAF. Since 2010 CCM Estimation chose to break the matches to "deletes", when the AFAQ evaluation matched a P-sample housing unit to a housing unit within the search area that was deleted in the block cluster, these matches have been broken. These cases are considered out of scope for this evaluation as well.

5. Results

All of the research questions require that the in-scope cases first be identified. Table 2 shows the missing housing units, and their status in terms of whether or not the units were matched in the AFAQ evaluation, and whether the match was to a "delete" in the 2010 Census. This information comes from the AFAQ evaluation data merged with the P-sample housing unit data, and the 2010 CUF. Only the 135 housing units that matched to valid housing units are of interest for the remainder of the evaluation.

Table 2: Status of the Missing Housing Units

Status	Count
Found in AFAQ Evaluation Matching	150
Matched to a Census "Delete"	15
Matched to a Valid Housing Unit	135
Not found on MAF, not in 2010 Census	4,040
Unresolved	3
Total Missing Housing Units	4,193

Data Source: AFAQ Results and the 2010 CUF

The first two research questions are relevant to housing unit matching outcomes, and are answered using the results from the AFAQ evaluation matching to the expanded address list.

1) The CCM search area consists of the sample block cluster plus the first ring of surrounding blocks adjacent to the block cluster. How many housing units on the final 2010 Census address list were geocoded in error to a block outside of the CCM search area?

2) How many CCM P-sample housing units were correctly coded as "missing" from the 2010 Census address list, but found on the MAF?

The first two research questions are answered by Table 3. Although the AFAQ evaluation did match within a five kilometer radius of the sample block cluster, none of the 135 housing units matched beyond the three kilometer ring.

Table 3: Source and Geography of Housing Units Found in AFAQ Matching[1]

Source and Geography	Unweighted Count of Housing Units	Weighted estimate of P-sample People in the Housing Units
Found in 2010 Census	**100**	**122,416**
In the CCM Search Area	24	33,399
In one kilometer ring, outside CCM Search Area	71	76,609
In three kilometer ring, outside one kilometer ring	5	12,408
Found on MAF, not in 2010 Census	**35**	**59,509**
In the CCM Search Area	34	56,488
In one kilometer ring, outside CCM Search Area	1	3,021
In three kilometer ring, outside one kilometer ring		
Total	135	181,926

1. Excludes housing units matched to census "deletes"
2. Blank cells indicate that there were no housing units with the combination of source and geography.

Data Source: AFAQ Results and the 2010 CUF

Of the 135 housing units that were matched, 100 were found on the 2010 Census address list. Of these, 76 were outside the CCM search area. There were 24 housing units that matched within

the CCM search area to the 2010 Census address list. Although these 24 housing units show differences between the CCM and AFAQ evaluation matching, this set of housing units are not indicative of the types of errors in the 2010 Census address list that are being studied in this report. As such, they will not be discussed further. The remaining 35 housing units were found on the MAF, but not in the 2010 Census. Almost all of these housing units were in the CCM search area, in fact, in the block cluster. Only one housing unit was found on the MAF, outside the CCM search area.

The third research question uses the weighted estimate of P-Sample people associated with the housing units CCM coded as "missing" that the AFAQ evaluation matched.

3) What impact would fixing the errors in the 2010 Census address list and in matching have on the CCM estimates of coverage for the household population?

The baseline one-cell DSE is 297,105,140. The baseline percent net undercount is -1.21 percent, and the baseline omissions estimate is 22,905,083. These estimates use the baseline values of DD, CE, P, M, E, and Census:

$$
\begin{aligned}
DD &= 294{,}709{,}980 \\
CE &= 274{,}200{,}057 \\
E &= 298{,}652{,}586 \\
P &= 286{,}121{,}350 \\
M &= 260{,}577{,}075 \\
Census &= 300{,}703{,}438
\end{aligned}
$$

3a. What if the CCM search area had been extended from the block cluster and surrounding ring of blocks to all blocks within three kilometers of the sample block cluster?

There are 163 unweighted P-sample people, representing 89,017 weighted persons in the 76 housing units that were matched to 2010 Census addresses outside of the CCM search area, but within three kilometers of the sample block cluster. If the search area is extended to blocks within three kilometers of the block cluster, then the definition of "match" and "correct enumeration" are modified so that any housing unit located in the newly defined search area is considered a match. As a result, persons in the newly matched housing units are considered matched, and it is assumed that a similar number of correct enumerations will be found in the E sample by extending the search area. As a result, these person enumerations are added to the estimates of matches and correct enumerations.

$$DSE_{alt\_1} = DD \times \frac{CE+89{,}017}{E} \times \frac{P}{M+89{,}017} = 297{,}100{,}099 \qquad (Equation\ 6)$$

$$pct\_under_{alt\_1} = \frac{DSE_{alt_1} - Census}{DSE_{alt\_1}} \times 100 = -1.21\% \qquad (Equation\ 7)$$

$$omiss_{alt\_1} = DSE_{alt\_1} - CE + 89{,}017 = 22{,}811{,}025 \qquad (Equation\ 8)$$

The dual system estimate that considers people in housing units matched in the extended search area is very close to the baseline estimate. The change to the estimate of matches is balanced by the change to the estimate of CEs. The resulting percent undercount shows essentially no change from the baseline measure. The estimate of omissions shows a reduction resulting mostly from the increase in the estimate of CEs.

3b. What if the 2010 Census address list had included the additional addresses that the missing-coded housing units matched to on the MAF?

There are two alternative CCM design options to consider. Either CCM could maintain its current search area of the block cluster and surrounding ring of blocks, or it could expand the search area to include everything within three kilometers of the block cluster.

Either way, if the 2010 Census address list had been constructed so that the additional MAF housing units were included, then the 2010 Census count, the data-defined census total, and E sample total would all include the people in these housing units. In addition, the number of matches and CEs would also increase by the number of P-sample people in these housing units.

(i) If CCM keeps its search area as the block cluster and first ring of surrounding blocks.

There are 71 unweighted P-sample people representing 56,488 weighted persons in the 34 housing units that were matched to the MAF, not in the 2010 Census, within the block cluster and surrounding ring of blocks.

$$DSE_{alt\_2a} = \text{DD} + 56,488 \times \frac{\text{CE}+56,488}{\text{E}+56,488} \times \frac{\text{P}}{\text{M}+56,488} = 297,102,692 \quad (Equation\ 9)$$

$$pct\_under_{alt\_2a} = \frac{DSE_{alt_{2a}} - (Census + 56,488\ )}{DSE_{alt\_2a}} \times 100 = -1.23\% \quad (Equation\ 10)$$

$$omiss_{alt\_2a} = DSE_{alt\_2a} - \text{CE} + 56,488 = 22,846,147 \quad (Equation\ 11)$$

(ii) If CCM expands the search area to include all blocks within a three kilometer radius of the block cluster. In this case, consider all housing units that matched to the MAF to be in the 2010 Census, in the E sample as CEs, and to be added to the total of matches.

There are a total of 74 unweighted P-sample people representing 59,509 weighted persons in the 35 housing units that matched to the MAF addresses within the three kilometer radius of the newly defined CCM search area.

To fully consider the implications of expanding the CCM search area to all blocks within the three kilometer ring, also consider people in the housing units that were matched to addresses in the 2010 Census, outside of the CCM search area. These people are matched and balanced by CEs in the E sample corresponding to the new

search area. The weighted P-sample people associated with these housing units are added to the match and CE rates. As a reminder, there are 89,017 weighted P-sample people who are in housing units matched to 2010 Census addresses outside of the CCM search area.

$$DSE_{alt\_2b} = \ DD + 59,059 \ \times \frac{CE + 59,059 + 89,017}{E + 59,059} \times \frac{P}{M + 59,059 + 89,017} = \ 297,097,542 \qquad (Equation\ 12)$$

$$pct\_under_{alt\_2b} = \frac{DSE_{alt_{2b}} - (Census + 59,059)}{DSE_{alt\_2b}} \times 100 = -1.23\% \qquad (Equation\ 13)$$

$$omiss_{alt\_2b} = DSE_{alt\_2b} - \ CE + 59,059 + 89,017 \ = \ 22,749,409 \qquad (Equation\ 14)$$

The largest difference between the baseline DSE and any of the alternative DSEs is a 7,598 reduction for DSE_{alt_2b}. Correcting the errors in the 2010 Census address list consistently decreased the DSE from the baseline estimate. The percent net undercount estimates either increased in absolute value or stayed the same, compared to the baseline percent net undercount. The estimate of omissions decreased by a maximum of about 150,000 people. Table 4 summarizes the baseline and alternative DSEs along with the corresponding percent net undercounts and omissions for each.

Table 4: Estimates of the DSE and Associated Coverage Measures for the Household Population Accounting for Errors in the Address List

Measure	DSE	Difference of DSE from Baseline	Percent Difference of DSE from Baseline	Percent Net Undercount	Omissions
Baseline	297,105,140	0	0	-1.21	22,905,083
Alt_1 Search *outside* the CCM search Area, including housing units on the 2010 Census address list	297,100,099	-5,041	-0.00170	-1.21	22,811,025
Alt_2a Search *inside* the CCM search area, expanding the address list to include housing units on MAF, not on 2010 Census address list	297,102,692	-2,448	-0.00082	-1.23	22,846,147
Alt_2b Search *outside* CCM search area, expanding the address list to include housing units on MAF, not on 2010 Census address list	297,097,542	-7,598	-0.00256	-1.23	22,749,409

Data Source: Calculations performed for this evaluation

6. Related Evaluations, Experiments, and Assessments

This evaluation is related to the 2010 Census Address Frame Accuracy and Quality Evaluation.

7. Conclusions and Recommendations

Expanding the address list used in matching to include housing units on the MAF in addition to housing units within a three kilometer radius of the sample block clusters did not result in much change to the housing unit matching outcomes or person DSEs. Expanding the area from three kilometer radius to a five kilometer radius resulted in no change. The majority of housing units that CCM classified as missing were not found on either the 2010 Census address list or the MAF.

Of the 135 housing units that the AFAQ evaluation matched either to "non-deletes" in the 2010 Census or to the MAF, 100 of the housing units were on the 2010 Census address list, while only 35 were on the MAF and missing from the 2010 Census address list. Only one of the 135 housing units was both located on the MAF but not the 2010 Census and found outside of the CCM search area. The matches to 2010 Census addresses outside of the CCM search area show some evidence for marginal gains in precision if CCM were to expand its search area to the one kilometer ring, with 71 housing units matching to 2010 Census addresses outside of the CCM search area, but within a one kilometer radius of the sample block cluster.

The hypothetical DSE estimates show that, under some simplifying assumptions, the DSE estimates for the household population are not affected much by the additional matched housing units. The largest difference between the baseline, one-cell DSE and the alternative DSEs is a reduction of less than 8,000 people, or 0.003 percent of the baseline estimate. The alternative DSEs are consistently lower than the baseline estimate.

The additional coverage measures computed from the alternative DSEs show that there is little change to the percent net undercount that would result from expanding the address list to include units within a three kilometer radius of the block cluster and units that were on the MAF but not in the 2010 Census. There is a maximum increase in the percent net overcount of 0.02 percent. Since all of the alternative DSEs are lower than the baseline estimate, and the correct enumeration estimates only increase corresponding to each alterative, the resulting estimates of omissions are also lower than the baseline estimate. The maximum change to the estimate of omissions was a decrease of about 150,000 people, and resulted from a smaller DSE and a larger estimate of CEs.

Any gains from expanding the CCM search area, geographically, in terms of the person estimation or housing unit estimation, would be small. These gains would be seen mostly in a decrease in the estimates of omissions, through an increase in the estimate of CEs.

Based on the evidence that refining the address list to include additional units on the MAF or expanding the search area geographically would result in little change to CCM estimates of coverage, we recommend no changes to the current search area or address list formation rules for CCM estimation purposes.

8. References

Census (2004). "Accuracy and Coverage Evaluation of Census 2000: Design and Methodology." U.S. Census Bureau. Washington, DC.

Johnson, N. (2010). "Study Plan for the Evaluation of Address Frame Accuracy and Quality." 2010 Census Planning Memoranda Series No. 146.

Johnson, N. and K. Kephart. "2010 Census Program for Evaluations and Experiments Report: Evaluation of Address Frame Accuracy and Quality." 2010 Census Planning Memoranda Series, forthcoming.

Mule, T. (2008). "2010 Census Coverage Measurement Estimation Methodology." DSSD 2010 Census Coverage Measurement Memorandum Series #E-18.

------ (2012). "2010 Census Coverage Measurement Estimation Report: Summary of Estimates of Coverage for Persons in the United States." DSSD 2010 Census Coverage Measurement Memorandum Series #G-01.

Mule, T. and S. Konicki (2012). "2010 Census Coverage Measurement Estimation Report: Summary of Estimates of Coverage for Housing Units in the United States." DSSD 2010 Census Coverage Measurement Memorandum Series #G-02.

Mulry, M., M. Moran, and P. Gbur (2011). "Study Plan for the Evaluation to Assess Effect of Census Coverage Measurement (CCM) Search Area and Census Address list Formation Rules on CCM Estimates." 2010 Census Planning Memoranda Series No. 161.

Olson, D. and R. Sands (2012). "2010 Census Coverage Measurement Estimation Report: Net Coverage Comparison with Post-stratification." DSSD 2010 Census Coverage Measurement Memorandum Series #G-12.

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

Seiss, M, et. al. (2011). "2010 Census Coverage Measurement: Status Imputation and Weight Trimming Software Requirement Specification." DSSD 2010 Census Coverage Measurement Memorandum Series #E-37R6.