U.S. Census Coverage Measurement Survey Plans

Vincent Thomas Mule, Jr.
U.S. Census Bureau

Abstract: This paper discusses the estimation plans of evaluating the coverage of the 2010 U.S. Census based on the results of the coverage survey. The U.S. Census Bureau will continue to estimate the net error of undercount or overcount based on the dual system estimation methodology. For 2010, the survey evaluation has been given the new objective of estimating the components of census coverage that include erroneous enumerations and omissions. This paper will provide more details on the estimation methodology being developed to produce these results.

1. Introduction

The purpose of the 2010 Census Coverage Measurement (CCM) program is the survey-based program to evaluate the coverage of the 2010 Census in order to improve future censuses, including 2020 and beyond. The CCM is designed to measure the coverage of housing units and the household population excluding group quarters and persons residing in group quarters. The CCM will provide estimates of the net coverage error and the components of census coverage by using a post-enumeration survey. Since the CCM is an evaluation, its results will not affect the 2010 Census counts. The other principal method to measure coverage is Demographic Analysis. See Devine and West (2010) for more information on the Demographic Analysis evaluation plans for the 2010 Census.

The 2010 CCM sample design is a large complex survey of 170,000 housing units in the United States (excluding remote Alaska) and 7,500 housing units in Puerto Rico. In the CCM survey, we conduct an independent enumeration of housing units and persons in housing units. The results are matched to census enumerations to identify coverage results. The CCM consists of five sampling activities, five data collection activities, and three matching activities prior to the estimation of census coverage. A high-level overview that shows the relationship and timing of the major CCM activities can be found in Whitford (2008). The CCM program will produce estimates for the United States and Puerto Rico. This paper focuses on the estimation methodology for producing the United States results.

Section 2 documents the coverage universes that the Census Coverage Measurement program will evaluate. Section 3 provides details on the new objective of estimating the components of census coverage, including erroneous enumerations and omissions. Section 4 gives background on dual system estimation used to estimate net coverage error focusing on the estimation for the household population. Section 5 provides a general approach to the estimation of net census coverage error and the components of census coverage. Section 6 describes an example figure of how the net coverage and components of census coverage are related. Section 7 provides some general information related to measuring the uncertainty of the estimates. Section 8 lays out

some of the proposed estimation domains for the CCM program. Section 9 lists the planned release dates for the net coverage and components of census coverage results.

2. Coverage Universes

Table 1 lists the two universes for coverage evaluation, and the several types of persons and living quarters that are out of scope for the CCM. All proposed CCM coverage estimates shown in this document apply only to the two universes that are in scope.

Table 1: CCM Evaluation Universes

Coverage Universes	Out of scope for CCM
1. People in housing units 2. Housing units	1. People in Group Quarters facilities 2. Group Quarter facilities 3. People and housing units in Remote Alaska 4. People experiencing homelessness or other transient living conditions

3. Components of Census Coverage

This section shows important aspects of how the CCM program will produce estimates of the components of census coverage. The implementation decisions are a trade off between the goals of the CCM program and the sample size, field work, and matching resources available. In order to implement Nonsampling Error initiatives to reduce potential measurement error in the estimates, the CCM program reduced the sample size from 300,000 to 170,000 housing units. The implementation of the net error concepts is similar to what we have done in the past. Some differences between net error and component error concepts are noted below.

Table 2 identifies the components of census coverage for the 2010 CCM for both people in housing units and housing units. For both universes, we will estimate the number of correct enumerations, erroneous enumerations, and omissions. Sections 3.1 to 3.4 provide more information on the components of census coverage.

Table 2: Components of Census Coverage

People in Housing Units	Housing Units
1. Correct enumerations 2. Erroneous enumerations 3. Whole-person census imputations 4. Omissions	1. Correct enumerations 2. Erroneous enumerations 3. Omissions

3.1. Correct Enumerations for Components

In the CCM, we will evaluate if the data-defined enumerations in the census are correct enumerations. An enumeration is considered to be correctly enumerated if the record corresponds to a person or housing unit that should have been included in the correct geographic area. Since we are producing national, state, county, and place estimates, the

definition of the correct geographic area will change depending on the area being evaluated.

For national-level estimates, an enumeration is considered to be correctly enumerated if the record corresponds to a person or housing unit that should have been included anywhere in the U.S. in the coverage universe. This criterion will apply to the estimates of the total population and other domains like demographic characteristics and census operational areas. For state, county, and place estimates, the definition will change to require that the person should have been enumerated in that area. When duplication or multiple inclusion of a person occurs, one of the enumerations will be determined to be where the person should have been counted on Census Day according to the residence rules. This enumeration will be classified as correct and the other enumeration(s) will be classified as the erroneous enumeration(s).

This definition of correct enumeration for components of census coverage is different than the definition of correct enumeration used by the CCM for estimating net error. The definition for net error is stricter, as it applies additional criteria to minimize the bias in our dual system estimates: the record must (1) have sufficient identification information, and (2) be enumerated in the specific geographic area referred to as the block cluster search area[1]. For our component estimation, we have decided to use a different definition that is more suitable for national, state, county, and place estimates. Using the net error definition produces an underestimate of the correct enumerations and an overestimate of erroneous enumerations in the census, based on using the stricter criteria of being included a smaller geographic area and having to report a full name and two characteristics.

3.2 Erroneous Enumerations for Components

For component estimation, the CCM program will report the number of erroneous enumerations. We will also estimate erroneous enumerations by type. When examining the reasons that a case can be erroneous, we will report the results by three groupings. The two groupings apply to people in housing units and to housing units:

- Persons or units that should not have been enumerated at all
- Duplicate person or housing unit enumerations
- Enumerations included in wrong location

Table 3 shows examples of several types of erroneous enumerations that are combined into one category of "should not have been enumerated at all." Some of the reasons for this combination are the following:

- Some of the situations in which cases should not be enumerated at all happen so infrequently that there is not a unique match code for each of them. Some examples are people who should have been enumerated in a group quarters, those born after Census Day, and those who died before Census Day. To minimize matching error, we do not want to implement clerical coding

[1] The geographic search area is the block cluster and the one ring of surrounding census collection blocks. A block cluster is one or more contiguous collection blocks, and averages 30 housing units.

procedures for infrequently occurring events. These reasons are identified by an overall match code of being erroneously included since they are all reasons for being out of scope for the CCM.

- A person may have been enumerated in a group quarters and also erroneously enumerated in a housing unit. The housing unit enumeration will be classified as an erroneous enumeration due to duplication in our results. This decision was made because the Further Study of Person Duplication showed that the 5.8 million estimated duplicates in the 2000 Census included 600,000 duplicates between people in housing units and group quarters (Mule, 2002).
- For housing units, the 2000 Housing Unit Coverage Study (HUCS) showed that, based on our proposed definition of correct, the two main causes of erroneous housing unit enumeration were duplication and not being a housing unit (Barrett et al., 2001).

Table 3: Types of Erroneous Enumerations (Should Not Have Been Enumerated at All) that Will Be Grouped Together

People in Housing Units	Housing Units
Fictitious Person should have been enumerated in Group Quarters Born after Census Day Died before Census Day Other reasons	Not a housing unit Does not exist Other reasons

For national-level estimates that include total population or demographic groupings like Hispanic origin population, there will be no erroneous enumerations due to being included in the wrong location. For the evaluation of states, counties and places, we will apply the requirement for being a correct enumeration that the person should have been counted in the geographic area being evaluated. Any case that should not be included in that area will be classified as an erroneous enumeration. Where possible for counties and places, we will break erroneous enumerations due to wrong location into more detail about where the person should have been enumerated.

3.3 Whole-Person Census Imputations

For people in housing units, we will tally and report the number of whole-person census imputations. While the goals and objectives of our program have been expanded, the CCM program has not been given the goal to evaluate the census imputation process. The CCM program cannot assess whether the individual whole-person census imputations are correct or erroneous because there is no way to follow up on records whose information is based solely on imputed values. Even though the CCM could determine the number of people who should have been enumerated at a particular housing unit, this would not be particularly useful in determining how census operations can be improved in the future. Since it is not in line with our goal of estimating components of coverage that include erroneous enumerations and omissions, the CCM will report the total number of whole-person imputations. Table 4 documents the five groupings of imputations that the CCM will use when reporting more information on census imputations. At the national level, these five groupings will be reported. This

may be an area where more can be done for the 2020 coverage evaluation based on the 2010 experience.

Table 4: Whole-Person Census Imputations by Category

Count Imputation 1. Status Imputation - No information about the housing unit; imputed as occupied, vacant, or non-existent. Those imputed as non-existent are removed from the census files. 2. Occupancy Imputation - Existence of housing unit confirmed, but no information as to occupancy status; imputed as occupied or vacant. 3. Household Size Imputation - Occupied status confirmed, but no information as to household count; the household count is imputed.
Whole-person Characteristic Imputation 4. Whole House Substitution - Population count known; all characteristics imputed for the entire household. 5. Total Allocated - Population count known; all characteristics imputed for some, but not all, persons in the household.

Note: Any housing unit imputed as occupied during count imputation will also have its household count imputed, which results in whole-person imputations.

For the housing-unit universe, there is not a category of housing-unit imputations. There are some census housing units that will require status imputation as defined in Table 4 since no information about the unit could be obtained. The CCM will evaluate all of the housing units in sample, including those that required status imputation, as to whether they were correct or erroneous. Because of this, it is not necessary to report housing-unit status count imputations separately for this universe.

3.4 Omissions

The CCM program will estimate the total number of omissions in the two coverage universes. At the national level for people in housing units, we will estimate the number of omissions by whether or not the housing unit was included in the census. There is no direct estimation methodology available to measure omissions. The estimation strategy is to use the results of the dual system estimation and the correct enumeration component to obtain an estimate of omissions. Section 5.2 provides more details on this estimation approach.

One group of omissions that will not be included in our analysis are those people or housing units removed during the census processing. Examples of these removals include a person deleted because of identified duplication in the census or housing units identified as non-existent during a field operation. Some of these removals may be in error and could lead to possible omissions in the Census. This area of evaluation was discussed during CCM planning but was dropped as the CCM program has expanded considerably, leading to resource concerns. This may be an area where more work can be done for components in the 2020 evaluation.

4. Dual System Estimation

4.1. Dual System Estimation Model

We will be implementing dual system estimation to estimate the true population of household population and housing units. The section will highlight the household poulation estimation. There is a long history of using dual system estimation in measuring coverage errors in a census (Wolter, 1986; Hogan 1993; Hogan 2003; U.S. Census Bureau 2004).

The standard Petersen (1896) or Sekar-Deming estimator assuming independence is expressed as:

$$N_{++} = N_{+1}\left(N_{1+}/N_{11}\right) \quad (1)$$

Where N_{++} is the total number of people,
N_{+1} is the number of people counted in the census
N_{1+} is the number of people counted in the survey
N_{11} is the number of people counted in both the census and the survey

4.2 Determining Correct Enumerations for Dual System Estimation

This first step in operationalizing the dual system estimator is to define and estimate the list or set of individuals "correctly" in the census. The determination of the correct enumeration status is done since the general net coverage model assumes that spurious events like duplications, nonexistent cases or out-of-scope cases, such as an individual born after Census Day, have been identified in both systems and accounted for in the estimation (Wolter 1986). In this context for dual system estimation, "correctly" has four dimensions:

1. Appropriateness
2. Uniqueness
3. Completeness
4. Geographic correctness

"Appropriateness" means that the person should be included in the census. People who die before or who were born after the census reference date (April 1 in the U.S.) are not part of the population (universe) to be measured. Similarly, records that refer to fictitious "people," tourists or animals are out-of-scope. "Uniqueness" refers to the fact that we wish to measure the number of people included in the census, not the number of census records. If more than one record refers to a single person, the count of records must be reduced for purposes of the dual system estimation. "Completeness" means that the census record must be sufficient to identify a single person. If it lacks sufficient identifying information, we cannot determine for all of those records whether the person was appropriately and uniquely included in the census, nor can we determine whether he or she was also included in the survey. Accurate matching or additional interviewing is

not possible for all of these cases[2]. For the dual system estimation, the definition for "sufficient information for DSE processing" is complete name and two characteristics. "Geographic correctness" means that people are included in the census where they should be included. Enumerations outside the defined search area that are counted in the census are not considered correctly included in the census for dual system estimation.

To estimate the number of people correctly included in the census, one must take a sample of all data-defined census enumerations. A sample of census person and housing unit records, referred to as the Enumeration (E) sample, are taken to determine how many of them were correctly counted in the right search area. To maximize correlation with the Population (or P) sample, the CCM first defines a set of sample areas. The sample areas are the block clusters. The P sample is the second independent sample for the dual system estimation. If a block is selected for sample, all census records coded to that block, even incorrectly, fall into sample. If the block contains many census housing unit records, it may be subsampled.

4.3. Determining the Proportion of People Captured Correctly in the Census

Having defined the set of correctly enumerated people, the next step in dual system estimation is to estimate the census coverage rate N_{11}/N_{1+}. This rate is also called the match rate.

The CCM program will independently list the housing units in the sample of block clusters. A sample of these listed housing units will be selected and those will be the P-sample housing units. We will then conduct the CCM Person interview at those housing units. The interviewing will result in a list of people who should have been enumerated in a housing unit in the census. This list constitutes the population or P sample.

The CCM will then attempt to match the P-sample people to the census records. A case can only be matched if the census record is located in the correct search area. The correct search area is the block cluster and the surrounding ring of collection blocks where the person or housing unit should have been counted in the census according to the census residence rules. Dual system estimation assumes that the P-sample people and housing units can be matched correctly to the census results and this determination is made with no error (Wolter 1986).

Conceptually, estimating the match rate entails (1) taking a sample of people, (2) determining whether they should be enumerated in the census, and (3) determining whether they were, indeed, correctly enumerated, using the same definitions as were used to determine the correct enumerations in the previous section. If a representative sample can be drawn of people who should have been enumerated and if we can determine whether they actually were correctly enumerated, then the dual system

---

[2] During clerical matching to support the component estimation, we are able to determine for some of these insufficient cases whether they were correct or erroneous based on results of matching people collected during the CCM Person Interview to these cases. These clerical matching results are not used for dual system estimation so that we do not introduce matching error.

estimation will produce asymptotically unbiased estimates. If each step can be approximately correct, the results will approach an unbiased estimate.

The CCM guards against unnecessarily introducing operational dependence by applying the same requirements on P-sample records that were applied to determine the correct enumeration status of the E sample in the previous section. The P-sample records are processed on the same four dimensions listed in Section 4.2. Out of scope records are screened out. Occasionally, survey duplicates occur and these are removed. Finally if the survey interview does not meet minimal standards the case is converted to a nonresponse and is handled by missing data methods.

The matching and followup result in the ability to estimate the number of people correctly counted in the survey (N_{1+}) and the number of people correctly counted in both the census and the survey (N_{11}). It is not the two estimates that are of specific interest but the ratio of the two for the dual system estimation.

People who move between the census reference date and the time of the survey interview present a challenge for designing a dual system estimation for a census application. First, people who move are more likely to be missed by the census and by the survey. Secondly, if a person has a different residence at the time of the survey than he did at the time of the census, one must decide where to sample him.

The 2010 CCM is doing a variation of how movers were treated for the 1990 PES. In the 1990 PES, movers were sampled where they lived at the time of the survey interview (nonmovers and inmovers since Census Day). We then searched the census records at their April 1 residence according to Census residence rules. This is known as procedure B (Marks 1979). For the 2010 CCM, we are also including people who moved out of the housing unit since Census Day but who have no probability of selection for Interview Day. One example of this is a high school senior who is enumerated at home on April 1st and then resides in a college dorm in September. These people are being included in the P sample since they were living in a housing unit on Census Day. Since they have moved to a group quarters, they have no chance of being selected in sample since group quarters are out of scope for the CCM coverage evaluation. This treatment is being called “Procedure or PES-B+.”

5. General Estimation Approaches

This section documents the general approaches for estimating net error and components of census coverage items.

5.1 General Estimation Approach for Dual System Estimation and Net Error

This section provides the general estimation approach that the CCM program will be using for the household population. This section lays out our approach for the estimation of the household population and how it is used to generate net coverage error estimates. While the focus is on the household population estimation, we highlight some of the differences for our housing unit estimation.

The dual system estimation requires the assumption that census capture probabilities be independent for all individuals on a list (Wolter 1986). The independence assumption can fail either due to causal dependence between the Census enumeration and the coverage survey enumeration or due to heterogeneity in capture probabilities. Causal dependence occurs when the event of an individual's inclusion or exclusion from one system affects their probability of inclusion in the other system. However, even if causal independence is true for all individuals, the independence assumption can be violated by heterogeneity of inclusion probabilities across different individuals. If the census inclusion probabilities are not equal for all individuals, the heterogeneity results in correlation bias.

For the CCM, we will use logistic regression modeling instead of post-stratification. The logistic regression modeling allows us to reduce the correlation bias in our total population estimates without having to include unnecessary high-order interactions as when forming post-stratification cells. Not having unnecessary high order interactions allows us to include additional variables in the model that can potentially help us reduce synthetic error for national, state, county, place and other estimates.

Haberman, Jiang and Spencer (1997) researched using logistic regression models instead of post-stratification to generate dual system estimates of the 1990 PES. Their research showed that separate models could be fit to determine the correct enumeration status of the E sample and the match status of the P sample. The results of these models could be used to generate population estimates.

Griffin (2005) documented the plans for several dual system estimation alternatives using these logistic regression ideas. Each plan has advantages and disadvantages to the total population estimates and estimates for specified sub-populations. Two estimators involve summing the predictions over all of the census records. The difference between the two was how the data-defined status was handled. The census processing is able to determine this status for every census record. The estimator we chose uses a predicted rate of being data-defined based on a model while an alternative estimator replaces the predicted rate by an indicator value of the data-defined status of the census record in the estimation. For large estimation domains, both produce negligible differences. However, the alternative estimator has a potential weakness for any small estimation domains that have relatively small number of data-defined cases. This alternative estimator may produce estimates that are too low unless the match rate model contains a covariate that can balance for the small number of data-defined cases. Mulry et al. (2008) showed this result when using this alternative estimator to generate block cluster-level estimates for clusters with low data-defined rates that did not include a covariate to balance the match rate model.

The three outcomes to be modeled by logistic regression for the household population are:

1. Was a census case data-defined?
2. Was the data-defined E-sample case a correct enumeration in the correct location for net error estimation?
3. Did a P-sample case match to a census enumeration in the correct search area?

For each of the three logistic regression models, Table 5 summarizes the successful outcome of the dependent variable , data utilized and whether weights will be utilized when running these three logistic regression models. The data-defined regression will use all of the census persons in housing unit records in the regression while the other two will use sample data. This is similar to the post-stratification estimates for the 1990 PES and March 2001 A.C.E. estimates where the data-defined rate for each post-stratum was determined using the entire census data.

For the other two regressions, we will use the sampling weights in the correct enumeration and match status regressions for those sample data. When running these models, the independent variables will be main effects or interactions of those effects. Our current plan is to include the same main effects in each model but not necessarily have the same interactions in each model.

Table 5: Logistic Regression Model Information

Successful Outcome of Dependent Variable	Data Source	Weights Used in Regression
Data-Defined	All Census People in Housing Units	No
Correct Enumeration	E-sample People	Yes
Match	P-sample People	Yes

We will use the results of the three logistic regression models to make predictions of the probability of being a data-defined enumeration, probability of being a correct enumeration and probability of being matched shown in equation 2. We can make the predictions by using a) the values of the independent variables of the census case and b) the estimated parameters of the respective models shown above. The independent variables will be the same as those used in the logistic regression models that generated the estimated parameters.

As part of this estimation, we will implement operations to account for missing data and reduce the sampling and nonsampling errors in our estimates. This includes imputation of missing characteristics, imputation of unresolved statuses, weight adjustment for non-interviewed P-sample housing units.

Correlation bias exists whenever the probability that an individual is included in the census is not independent of the probability that the individual is included in the CCM or there is heterogeneity in the capture probabilities. This form of bias generally has a downward effect on estimates, because people missed in the census may be more likely to also be missed in the CCM.

We are planning on doing the adjustment by individual years of age to the 18+ Black and 18+ non-Black male populations. This is different than the adjustments in A.C.E. Revision II that used three age categories: 18-29, 30-49, and over 50. Shores (2003) documents the correlation bias adjustment for the A.C.E. Revision II estimates. Bell (1999) documents using the two group method and age groupings to adjust logistic regression dual system estimates of the 1990 PES data as part of the 2000 A.C.E. research. Females and children have a correlation bias adjustment factor equal to 1.

The correlation bias adjustment is done so that the final dual system estimation sex ratios based on the adjusted male estimate will agree with the Demographic Analysis-adjusted sex ratios for Blacks and non-Blacks. Since we are doing the adjustment for individual years of age, the Demographic Analysis and CCM estimates used in the calculations will be smoothed by using a 5 year rolling average.

For the 2010 CCM, we will use the following estimator for the household population

$$N_o = \sum_{j \in C} \pi_{dd,j} \frac{\pi_{ce,j}}{\pi_{m,j}} CB_j \quad (2)$$

Where $\pi_{dd,j}$ is the predicted probability census case j is data-defined,
$\pi_{ce,j}$ is the predicted probability that the census case is a correct enumeration
$\pi_{m,j}$ is the predicted probability that the census case was matched
j is the a census enumeration
CB_j is the correlation bias adjustment factor appropriate for enumeration j
C is the estimation domain (total, tenure status, etc.).

The CCM program will use a similar logistic regression approach to estimate the number of housing units. One major differences is that it only requires the logistic regression models for the correct enumeration and match statuses shown in Table 5. The second is that there will not be a correlation bias adjustment as shown in equation 2.

Estimates of net error will be computed based on the difference of the dual system estimate and the census count. A positive estimate shows an undercount and a negative estimate of net error shows an overcount.

$$NetError = DSE - Census \quad (3)$$

We will also continue to report the estimate of percent net undercount. The percent net undercount is the net error estimate calculated above divided by the dual system estimate expressed as a percentage.

$$Percent\ Undercount = \frac{DSE - Census}{DSE} \quad (4)$$

The CCM can generate dual system estimates for census operational areas, but not for those who should have been enumerated by the result of specific operations. For example, for operational areas like Mailout/Mailback TEAs, we can identify for both the census and the independent Population (P) sample the geographic area where the person should have been counted in the census. We can use the characteristics of those

geographic areas like Mailout/Mailback to generate a dual system estimates for those types of areas. We have not developed a methodology to estimate what should have been the true population of those who were enumerated by mail returns or those by nonresponse followup enumerations. Being enumerated by a mail return is a characteristic of the census enumeration that serves as the first system of the dual system estimation. To maintain independence when interviewing the P sample, we do not ask the people if and how they were enumerated during the census. Based on the data available, several assumptions would need to be made to generate these types of estimates. There are no plans to produce these estimates in our production estimates. This is an area that may be expanded in future census coverage surveys.

5.2 General Estimation Approach for Components of Census Coverage

The general estimation approach for components falls into three areas. First, the number of whole-person census imputations for people in housing units will be tallied from the Census files. Second, the estimates of correct and erroneous enumerations will be design-based estimates using the matching, followup, and processing results of the person and housing unit cases from the CCM sample of census enumerations. To control variance, we will implement a two-stage ratio adjustment during the estimation by taking advantage of the finite population total of census enumerations. We will also implement operations to account for missing or unresolved enumeration status and missing characteristics.

The third is the estimation of omissions. Experience from the 1950 Post-Enumeration Survey showed that applying data from a post-enumeration survey to estimate the total population by estimating directly the number of omissions and erroneous enumerations in the census resulted in a measured undercount that was well below the results produced by Demographic Analysis (U.S. Census Bureau 1960). To account for this underestimation of omissions in the total population, the Census Bureau has used dual system estimation based on the survey data to estimate the total population.

Since we use dual system estimation, there is an issue trying to estimate omissions. For practical reasons, the procedure is implemented using a very restrictive application of "completeness" and "correct location" as part of the definition of correct enumeration. This leads to an overstatement of both erroneous enumerations and nonmatches in a way that should balance for net error estimation, but that does not work for omissions. The only way, conceptually, to directly address this would be to apply matching and followup to the whole country in the credible way. Since we will rely on dual system estimation for estimating the true population for net error estimation, we will use this estimate in determining the population that was missed.

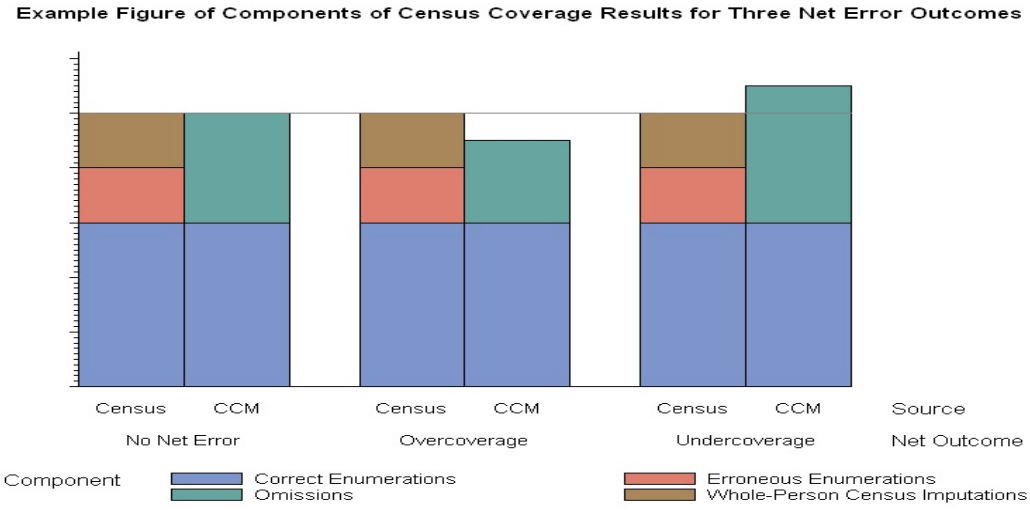
We will estimate omissions by subtracting the estimate of correct enumerations from the dual system estimate as shown in the formula below. Omissions are people who were either a) not enumerated in the Census; or b) enumerated based on a reported population count for the housing unit, but enough information wasn't collected for the individual persons to allow the CCM program to determine if they were correctly or erroneously included. This second group are people in housing units that the census enumerated but where more information about the people could have been obtained.

Since our estimator for omissions uses the dual system estimate, we will not produce estimates of omissions for any estimation domain where we do not produce dual system estimates. Because of this, we will not produce estimates of omissions for the results of census operations.

$$Omissions = DSE - Correct\ Enumerations \quad (5)$$

6. Relationship Between Net Coverage and Components of Census Coverage

The figure below shows the relationship between the estimates of net coverage and the components of census coverage. The figure shows the relationship for three net coverage error outcomes. This figure does not contain any real data and the sizes of any bars are for illustrative purposes.



The left part of figure shows the relationship when there is no net coverage error. The census and the CCM have estimated the same population total. The left bar shows the census count broken into the three components of correct enumerations, erroneous enuemrations and whole-person census imputations. The CCM bar is divided into the estimate of correct enumerations and the omissions. The esimate of correct enumerations is based on the E-sample so it is the same in each column. In this example with no net coverage error, we see that the erroneous enumerations and whole-person census imputations balance out the estimate of omissions.

The middle part of the figure shows the relationship between net coverage and components when the census has overcoverage. The CCM has estimated a lower population than the census count reflected in the height of those bars. The census count is again broken into the three components of correct enumerations, erroneous enumerations and whole-person census imputations. The CCM estimate is broken into the correct enumerations and omissions. One thing to point out about this possible scenario is that omissions is a component of coverage and *not* a net result. If you were to take the estimate of omissions and add that to the census count, you would be moving the census result further away from the CCM estimate. The right part of the figure shows

the relationship when the census has undercoverage. The CCM has estimated a higher population than the census count.

7. Measures of Uncertainty

All estimates will include appropriate measures of uncertainty. This will include the standard error of the estimates due to sampling error. They will be produced using a random group jackknife variance methodology.

There are currently no plans to include other sources of uncertainty in the estimates. These include the increase in variance for unresolved status imputation. We are exploring the possibility of producing estimates of mean squared error of the estimates. These estimates of error would account for the synthetic bias in the estimates in addition to the sampling variance.

8. Estimation Domains

In order to evaluate the 2010 Census and help with the planning for the 2020 Census, the CCM program will be producing coverage results for the following domains for the household and housing unit poulations::

- National-level
- Demographic or Housing Unit characteristics
- Census operational areas including Type of Enumeration Areas
- Census operations like Mailout/Mailback or Nonresponse Folllowup returns

Table 6 shows the net coverage and components of census coverage results that will be produced as well for states, counties and places. Based on feedback from outside advisory groups, the CCM program evaluated what state, county and place estimates could be produced. That evaluation led to the determination that we could produce both net and component estimates for the 50 states and the District of Columbia. For net coverage error, we determined that we could produce estimates for individual counties or places with over 100,000 people. For the remaining counties or places in a state, a balance of state estimate would be produced as well. For components of census coverage, we determined that we could produced estiamtes of components for individual counties or places with over 500,000 people. Again for the remaining counties or places, a balance of state estimate would be produced.

Table 6: Estimation Domains for Governmental Entities

Governmental Entity	Net Coverage Error	Components of Census Coverage
States	50 States and DC	50 States and DC
Counties and Places	100,000+ Balance of State for remainder	500,000+ Balance of State for remainder

The different threshold of 500,000 for components as compared to 100,000 for net error is based on the fact that the estimator for correct and erroneous enumerations is a design-

based ratio-adjusted estimator. There are some counties with over 100,000 people that do not have sample. In looking at the projected sample sizes, we determined at this point that counties with over 500,000 people would have enough sample to support the release of estimates.

9. Release of Estimates

The CCM program will release the estimates of net coverage error in July 2012. In September 2012, the CCM program will release the components of census coverage results.

References

Barrett, D. F., Beaghen, M., and Smith, D. (2001), "ESCAP II: Census 2000 Housing Coverage Study," Executive Steering Committee For A.C.E. Policy II, Report No. 17., October 17, 2002.

Bell, W. (1999), "Accuracy and Coverage Evaluation Survey: Combining Logistic Regression DSEs (Target Model) with 1990 Demographic Analysis Results," DSSD Census 2000 Procedures and Operations Memorandum Series Q-12.

Devine, J. And West (2010), "U.S. Demographic Analysis Plans" paper presented at the Joint Statistical Meetings, Vancouver, Canada, August 2010.

Griffin (2005), "Net Error Estimation for the 2010 Census", DSSD 2010 Census Coverage Measurement Memorandum Series #2010-E-1.

Haberman, S. J., Jiang. W., and Spencer B. D. (1998), " Activity 7: Develop Methodology for Evaluating Model-Based Estimates of the Population size for States Final Report", prepared by NORC for the Bureau of the Census under contract no. 50-YABC-2-66023.

Hogan, H. (1993), "1990 Post-Enumeration Survey: Operation and Results," Journal of the American Statistical Association.

Hogan, H. (2003), "The Accuracy and Coverage Evaluation: Theory and Design." Survey Methodology, 29, 129-138.

Mule, T. (2002), "A.C.E. Revision II Results: Further Study of Person Duplication" DSSD A.C.E. Revision II Estimates Memorandum Series # PP-51, U.S. Census Bureau report, December 31, 2002.

Mulry, M. H., Spencer, B. D., Mule, T., Nguyen, N. (2008), "Direct Estimates As a Diagnostic for Dual System Estimators Based on Logistic Regression". JSM *Proceedings.* Survey Research Methods Section. American Statistical Association, Alexandria, VA.

Petersen, C.G.J. (1896), The Yearly Immigration of Young Plaice into the Limfjord from the Grman Sea. Report of the Danish Biological Station. 6, 1-48.

Shores, R. (2003), "A.C.E. Revision II: Adjustment for Correlation Bias," DSSD A.C.E. Revision II Memorandum Series, #PP-53, Washington, D.C.

U.S. Bureau of the Census (1960), "The Post-Enumeration Survey: 1950," Technical Paper 4, Washington, D.C.

Whitford, D. (2008), "Overview of 2010 Census Coverage Measurement Program," DSSD 2010 CENSUS COVERAGE MEASUREMENT MEMORANDUM SERIES #A-19, July 9, 2008.

Wolter, K.M., (1986), "Some Coverage Error Models for Census Data" Journal of the American Statistical Association, 81, 338-353.