Framework for Census Coverage Error Components

Mary H. Mulry and Donna K. Kostanich[1]
U.S. Census Bureau, Washington, DC 20233

Abstract

A major goal and challenge for coverage measurement in 2010 is to design a survey that measures the components of coverage error, namely erroneous enumerations and omissions. The Census Bureau's previous coverage measurement surveys were designed primarily to estimate net census error using Dual System Estimation (DSE). To improve the accuracy of estimates of net error, the Census Bureau's DSE has relied on balancing some of the components of error, meaning some census omissions offset some erroneous inclusions in a manner that preserved the net error. As a result, the process produced inflated estimates of omissions and erroneous inclusions. This paper provides a framework for overcoming these inflated estimates of component errors. It also explicitly defines the individual components of error and how these components relate to traditional net error concepts.

Keywords: dual system estimation, undercount, omissions, erroneous enumerations, 2010 Census

1. Introduction

A major goal and challenge for coverage measurement in 2010 is to design a survey that measures the components of coverage error, namely erroneous enumerations and omissions (Kostanich, Whitford, and Bell 2004). Previous coverage measurement surveys, including the 2000 Accuracy and Coverage Evaluation (A.C.E.) (U.S. Census Bureau 2004) and the 1990 Post Enumeration Survey (PES) (Hogan 1992 1993), were designed primarily to estimate net census error using Dual System Estimation (DSE). To improve the accuracy of estimates of net error, the implementation of the DSE has relied on balancing some of the components of error, meaning some census omissions offset some erroneous inclusions in a manner that preserved the net error. Essentially this has entailed using a very strict definition for measuring correct enumerations (Hogan 2003). To be classified as a correct enumeration, the enumeration had to be included in the right location. Right location was defined as the block or surrounding ring of blocks, known as the search area. Additionally, only those enumerations with complete name and two characteristics were eligible for matching. The remaining enumerations which are referred to as those with insufficient information (ignoring census imputations) could not qualify as correct enumerations under this strict definition. These criteria have resulted in inflated estimates of omissions and erroneous inclusions.

This paper provides a framework for overcoming these inflated estimates of component errors. It also explicitly defines the individual components of error, how these components relate to traditional net error concepts, and how the various survey activities feed into measurement of these components.

2. Strategy

The strategy for measuring the components of coverage error essentially involves expanding the definition of what is a correct enumeration. (Note that the plans involve continuing to maintain the narrower definition for purposes of estimating net error.) First, an expansion of the definition of correct location permits the determination of whether an enumeration was included in the right county, state, or even just somewhere in the nation rather than limiting correctness to only those enumerations that are in the right small geographic area; i.e., search area. Accomplishing this will require collecting additional information on where an enumeration should have been included. This was not done for the 2000 A.C.E. or the 1990 PES. These surveys only collected information to determine if the enumeration was included in the right location (search area) and could therefore not identify where those enumerations in the wrong location should have been included. The A.C.E. and the PES treated these enumerations as omissions in the geography where they should have been enumerated and as erroneous enumerations in the geography where they were enumerated and were thus offsetting for net error calculations. Even after determining the location where an enumeration should be included, additional matching to that location is needed to determine if the census had included the enumeration twice.

[1] This report is released to inform interested parties and encourage discussion of work in progress. The views expressed on statistical, methodological, and operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.

Determining enumeration status for some records that don't have complete name and two characteristics requires an expansion of the definition of enumerations eligible for matching in order to measure component errors. The exact definition is still under consideration; however, in order to have confidence in identifying matches and nonmatches, it is likely that name will be required. The matching and followup will attempt to determine whether the enumerations with adequate information are correct or erroneous. For enumerations without enough information to match confidently, imputations or some assumptions will be needed. Previously for A.C.E. and PES, persons whose census enumerations had insufficient information were not eligible for matching. These enumerations would be represented as omissions if listed in the independent sample.

On the surface this strategy for estimating components of coverage error appears fairly straightforward. Consider a very simplistic approach. A better estimate of correct enumerations that also reflects expanded definitions of correct location for the level of geography desired can be subtracted from the census count to obtain a better estimate of erroneous enumerations. An estimate of omissions can then be obtained by subtracting the estimate of correct enumerations from the estimate of total population resulting from the DSE. Current plans include investigating more sophisticated estimation methods such as post-stratification, ratio adjustment, or other types of models.

An extremely important aspect of estimating the error components is to obtain good data that is consistent with the concepts being measured. Survey activities will include data collection, several types of matching, potential field followup, and probably most critical, data coding. Understanding how these data sources and activities relate to the specific concepts is one of our major challenges. Furthermore, lack of information or perhaps inconsistent information will impose limitations on what can be measured.

This document begins with a discussion of the concepts traditionally used for the DSE model. These concepts are viewed from a different perspective in order to identify important pieces of information that are not distinguishable in the DSE model. The balancing assumption required for the DSE is discussed. Then there is a discussion of the components of coverage error and missing pieces of information. The next section contains a description of the implementation of the DSE in practice. This is followed by an explanation of which survey operations will be used to measure the component parts. Finally, some ideas are presented on possible estimation approaches.

3. The Dual System Model for Net Error

The purpose of dual system estimation is to estimate the true population total so that net coverage error in the census can be estimated. The DSE model works by obtaining two independent measures of the true population and matching to measure those persons captured in both systems. One system is the census and the other is an independent enumeration of the population. This section reviews the DSE model assumptions and how these assumptions relate to our implementation of the DSE. Here the DSE model is viewed from an ideal perspective meaning that the truth is known about all census enumerations, the independent enumeration includes only enumerations eligible for inclusion in the census, and there are no matching errors.

For all census enumerations, the assumption is that their correct or erroneous enumeration status is known. This assumption alleviates the need for a discussion of imputation although in practice there are enumerations whose status is unknown. Furthermore, it is assumed that all erroneous enumerations are classified as being erroneous either because it is in the wrong location or because it should not have been included anywhere. The latter category include duplicates, fictitious, not a resident of the U.S., died before Census Day, and born after Census Day. For census enumerations included twice; i.e., duplicates, at most one of these enumerations can be a correct enumeration or an erroneous enumeration due to wrong location. In the cases where a person has three or more enumerations, only one can be correct. Even though the assumption is that the true status of all census enumerations is known, a separate category for some enumerations that are missing name and two characteristics is necessary to mirror the data limitations in practice.

For the sake of the derivation, the assumption is the performance of a completely independent enumeration of the entire country. This is referred to as the P-Census since there is no sampling involved. The DSE model does not require that the P-Census capture everyone that should have been enumerated in the census, but it does require the assumption that those who are included in the P-Census should have been included in the census. This assumption implies that if the P-census did have any false inclusions, they were detected and removed. P-Census are also assumed to be captured at the correct location and all P-Census have sufficient information for matching purposes.

Next assume the entire P-Census can be matched to all census enumerations with name and at least two characteristics. This subset of census enumerations with a name and atleast two characteristics will be referred to as the Matching Universe. A "match" can only represent an enumeration included at the correct location. Also assume

that matching across the whole country is possible. Also the identification of those census enumerations included in the wrong location is possible, but these will not be considered "matches" since this exercise examines assumptions in the definitions used for the traditional DSE. Another important assumption is no matching error.

Table 1 contains the status of population for the census and the matching universe crossed by the status for the P-Census. In Table 1, subscripts *i,j* are defined as follows:

$$i = \begin{cases} 1 & in \quad MatchingUniverse \\ 0 & otherwise \end{cases}$$

$$j = \begin{cases} 1 & in \quad P-Census \\ 0 & otherwise \end{cases}$$

Let:
CE represents correct census enumerations in correct location,
WL represents census enumerations in the wrong location
II represents census enumerations with insufficient information for matching
NDD represents non-data-defined census enumerations, those that do not have at least 2 characteristics reported.
EE represents erroneous census enumerations
OM represents census omissions

Note that II and NDD actually represent enumerations that should be included in the Census. This is also the case for WL under a broader definition of correct enumeration.

Note that $CE = CE_{11} + CE_{10}$
$WL = WL_{11} + WL_{10}$
$II = II_{01} + II_{00} + EE_{00\text{-}II}$
$NDD = NDD_{01} + NDD_{00} + EE_{00\text{-}NDD}$
$OM = OM_{01} + OM_{00}$

Also note that the total census count includes all correct and erroneous enumerations. This census count is given by:

Census =
$CE_{11} + CE_{10} + WL_{11} + WL_{10} + II_{01} + II_{00} + NDD_{01} + NDD_{00} + EE_{10} + EE_{00\text{-}II} + EE_{00\text{-}NDD}$.

The total true population includes all correct enumerations, erroneous enumerations in the wrong location, and omissions. This true total population is given by:

True Pop =
$CE_{11} + CE_{10} + WL_{11} + WL_{10} + II_{01} + II_{00} + NDD_{01} + NDD_{00} + OM_{01} + OM_{00}$.

The net error in the census is then given by:

NetCensusError = True Pop - Census
$= OM_{01} + OM_{00} - EE_{10} - EE_{00\text{-}II} - EE_{00\text{-}NDD}$

Also note that the P-Census count is given by:

P-Census = $CE_{11} + WL_{11} + II_{01} + NDD_{01} + OM_{01}$.

4. The DSE Model Assumptions

Dual system estimation (DSE) is used to estimate the total true population. This section starts out by looking at the expression for the DSE under this ideal scenario outlined in the previous section. Next is a discussion of the assumptions needed so that this DSE provides an unbiased estimate of the total (U.S.) population. Recall that a search of the entire census matching universe for enumerations in the P-Census is possible. Only correct enumerations in the correct location are considered "matches". Also the enumeration status and correct location for all census enumerations is known even when the enumeration is not in the matching universe. Also note that all duplicate census enumerations have been identified and their correct enumeration status is known. One form of the DSE is given by:

$$DSE = CE\,\frac{P}{M}$$

where:
CE is the number of correct census enumerations in the matching universe.
P is the number of enumerations in the P-Census.
M is the number of the P-Census matching to correct census enumerations in the matching universe that are in the correct location.

Therefore: $CE = CE_{11} + CE_{10}$
$P = CE_{11} + WL_{11} + II_{01} + NDD_{01} + OM_{01}$
$M = CE_{11}$

and the dual system estimate of the population can be written as:

$$DSE = (CE_{11} + CE_{10})\,\frac{(CE_{11} + WL_{11} + II_{01} + NDD_{01} + OM_{01})}{CE_{11}}.$$

Assumption 1

The basic assumption underlying this DSE is that the proportion of the total True Population correctly enumerated in the census equals the proportion of the P-Census enumerated in the census, which can be expressed as the proportion of the P-Census that would match if all Census enumerations were in the matching universe. This assumption holds when the census and P-Census are

independent and can be expressed as:

$$\frac{CE + WL + II_{01} + II_{00} + NDD_{01} + NDD_{00}}{TruePop} = \frac{CE_{11} + WL_{11} + II_{01} + NDD_{01}}{CE_{11} + WL_{11} + II_{01} + NDD_{01} + OM_{01}} \quad (1)$$

Using algebra, *TruePop* equals Equation (2) and is called the DSE:

$$TruePop = (CE + WL + II_{01} + II_{00} + NDD_{01} + NDD_{00}) \times \frac{CE_{11} + WL_{11} + II_{01} + NDD_{01} + OM_{01}}{CE_{11} + WL_{11} + II_{01} + NDD_{01}} \quad (2)$$

From a more practical perspective, the Census Bureau's implementation of the DSE has to deal with some census enumerations not having enough data to match with confidence and with not being able to search the entire census to determine whether a person is enumerated or whether an enumeration is correct or erroneous. The implementation of the DSE consists of matching enumerations in the P-Census to census enumerations in the matching universe. This matching is only done within the search area.

Assumption 2

The Bureau's implementation of the DSE assumes that correct enumerations in the matching universe are included in the P-Census at the same rate as all correct enumerations. This assumption can be expressed as:

$$\frac{CE_{11} + WL_{11}}{CE + WL} = \frac{CE_{11} + WL_{11} + II_{01} + NDD_{01}}{CE + WL + II_{01} + II_{00} + NDD_{01} + NDD_{00}} \quad . \quad (3)$$

Assumption 3

The search to determine whether an enumeration is correct or erroneous and whether a P-Census person does or does not match a census enumeration has to be limited to assure confidence in designating nonmatches as well as matches. The implementation is designed so that the proportion of enumerations representing people that should be enumerated but are called erroneous because they are in the wrong location equals the proportion of matches that are not found because they are in the wrong location and called nonmatches. This assumption is equivalent to saying the percentage of correct enumerations found because they are in the correct location equals the percentage of matches found because the enumerations are in the correct location, which can be written as follows:

$$\frac{(CE_{11} + CE_{10})}{(CE_{11} + CE_{10} + WL_{11} + WL_{10})} = \frac{CE_{11}}{CE_{11} + WL_{11}} \quad . \quad (4)$$

Therefore,

$$\frac{(CE_{11} + CE_{10})}{CE_{11}} = \frac{(CE_{11} + CE_{10} + WL_{11} + WL_{10})}{CE_{11} + WL_{11}} \quad . \quad (5)$$

Substituting for Equations (3) and (5) in Equation (2) for *TruePop* produces the Census Bureau's DSE:

$$DSE = (CE_{11} + CE_{10}) \frac{(CE_{11} + WL_{11} + II_{01} + NDD_{01} + OM_{01})}{CE_{11}} \quad (6)$$

Using algebra, Equation (6) can be rewritten as:

$$DSE = (CE_{11} + CE_{10} + WL_{11} + II_{01} + NDD_{01} + OM_{01}) + \frac{CE_{10}}{CE_{11}}(WL_{11} + II_{01} + NDD_{01} + OM_{01}) \quad (7)$$

The DSE will be equal to the true population (*TruePop* in Equation (2)) and provides an unbiased estimate of net census error if the last term on the right-hand side of Equation (7), called the fourth cell (C4), is equal to:

$C4 = WL_{10} + II_{00} + NDD_{00} + OM_{00}$.

This is in fact the case under the assumption that being correctly included in the census (at the correct location) and in the matching universe is independent of being in the P-Census as given by:

$$C4 = WL_{10} + II_{00} + NDD_{00} + OM_{00} = \frac{CE_{10}}{CE_{11}}(WL_{11} + II_{01} + NDD_{01} + OM_{01}) \quad (8)$$

When the underlying assumptions 1, 2, and 3 hold as reflected in Equations (1), (3), and (4) respectively, the DSE will provided an unbiased estimate of the true total population and the net census error, which gives:

$$DSE = CE_{11} + CE_{10} + WL_{11} + WL_{10} + II_{01} + II_{00} + NDD_{01} + NDD_{00} + OM_{01} + OM_{00} \quad . \quad (9)$$

DSE - Census = NetCensusError (10)
= $OM_{01} + OM_{00} - EE_{10} - EE_{00\text{-}II} - EE_{00\text{-}NDD}$.

5. Components of Census Error Definitions and Notation

Besides net census error, another interest is measuring the components of coverage error (meaning omissions and erroneous enumerations) at various levels of geography. To do this requires additional information that is not available from the implementation of the dual system model described above. This section defines the components of error that are of interest under the ideal scenario.

For estimating net error a very strict definition of correct enumeration is used. To be considered correct the census must have included the enumeration in the correct search area. For the components of error, the interest is in the measurement of a variety of situations such as whether the enumeration included in the right county, state, or even just included somewhere in the U.S. For purposes of this discussion, the focus is on the components of error for the whole U.S. Under the assumption that it is possible to determine the location where a person should have been included, the different definitions of correct enumeration are mainly a tabulation issue.

Therefore, the goal is to obtain estimates of:

Erroneous Enumerations = $EE_{10} + EE_{00\text{-}II} + EE_{00\text{-}NDD}$

Omissions = $OM_{01} + OM_{00}$.

5.1. Erroneous Enumerations

Note that subtracting the correct enumerations used for net error, ($CE = CE_{11} + CE_{01}$), from the census count does not give an unbiased estimate of erroneous enumerations:

Census - CE = (11)
$(WL_{00} + WL_{10}) + (II_{01} + II_{00}) + (NDD_{01} + NDD_{00}) + (EE_{10} + EE_{00\text{-}II} + EE_{00\text{-}NDD})$

To obtain an unbiased estimate of erroneous enumerations, additional information is needed to estimate:

$$EE_{10} + EE_{00\text{-}II} + EE_{00\text{-}NDD}. \quad (12)$$

Additional data are also needed to estimate the WL, II, and NDD terms in Equation (11).

5.2. Omissions

Note that using the nonmatches from the P-Census does not give an unbiased estimate of omissions:

$$P - M = WL_{11} + II_{01} + NDD_{01} + OM_{01} .$$

An expression for omissions follows from Equation (10) and is given by:

$OM_{10} + OM_{00}$ = (13)
NetCensusError + $(EE_{10} + EE_{00\text{-}II} + EE_{00\text{-}NDD})$

6. The DSE Model in Practice

This section summarizes how the DSE model is implemented in practice. Obviously, it is not practical to do an independent enumeration of the entire U.S. nor is it feasible to accurately match all independent enumerations against all census enumerations. Therefore the DSE model is implemented on a sample basis and the matching is restricted to a small geographic area referred to as the search area. To estimate the true total population using a sample of the census, called the E-sample, and a sample of the P-Census, called the P-sample, the DSE is written as:

$$\hat{DSE} = \hat{CE}\,\frac{\hat{P}}{\hat{M}}$$

where:

$\hat{CE}$ is the estimated number of correct census enumerations in the Matching Universe.

$\hat{P}$ is the estimated number of enumerations in the P-Census.

$\hat{M}$ is the estimated number of the P-Census matching to correct census enumerations in the matching universe within the search area.

The expected value of these terms are given by:

$$E[\hat{CE}] = CE_{11} + CE_{10}$$

$$E[\hat{P}] = CE_{11} + WL_{11} + II_{01} + NDD_{01} + OM_{01}$$

$$E[\hat{M}] = CE_{11} \quad .$$

The first stage of sampling consists of selecting a sample of small geographic areas referred to as block clusters. A simplifying assumption is that the E-Sample includes all census enumerations in the selected block clusters. (In reality, any census enumeration that is represented by an imputed person or non-data-defined record is excluded from the E-Sample.) From these same block clusters, an independent enumeration of the population is conducted which comprises the P-Sample. Therefore, the E-Sample is a sample of all census enumerations represented in the heavy line in Table 1, and the P-Sample is a sample of the

P-Census, the double-lined box in Table 1. Persons who have moved between Census Day and the day of P-sample enumeration, called the Person Interview Day, are treated differently in these samples. The E-Sample, by default, includes these movers at their Census Day residence. The P-Sample includes persons who have moved into the sampled block since Census Day if they had moved from another housing unit in the U.S.

The P-Sample is only matched to matchable census enumerations in the search area which are represented by the shaded area in Table 1. For persons who have moved into the sampled block since Census Day, the P-Sample is matched to the block they resided in on Census Day. The matches provide an estimate of correct enumerations in the correct location that were included in the P-Sample. Therefore, the P-Sample provides the estimates:

$$\hat{P} = \hat{M} + N\hat{M}$$

where M represents matches and NM represents nonmatches. The expected values are given by:

$$E\left[\hat{M}\right] = CE_{11} \text{ and}$$

$$E\left[N\hat{M}\right] = \mathrm{WL}_{11} + \mathrm{II}_{01} + \mathrm{NDD}_{01} + OM_{01} \ .$$

Note that it is not possible to distinguish which of the P-Sample nonmatches are correct enumerations not in the matching universe, enumerations treated as erroneous because they are in the wrong location or out of the matching universe, or omissions from the census.

The E-Sample provides the estimates:

$$\hat{E} = C\hat{E} + E\hat{E}$$

where CE represents correct enumeration in the correct location and EE represents erroneous enumerations that should not have been included anywhere or those enumerations that are in the wrong location.

The expected values are given by:

$$E\left[C\hat{E}\right] = CE_{11} + CE_{10}$$

$$E\left[E\hat{E}\right] = (\mathrm{WL}_{11} + \mathrm{WL}_{10}) + (\mathrm{II}_{01} + \mathrm{II}_{00}) + (\mathrm{EE}_{10} + \mathrm{EE}_{00\text{-II}})$$

$$E\left[\hat{E}\right] = E[C\hat{E}] + E[E\hat{E}] \ .$$

The estimate of CE reflects enumerations included (CE_{11}) and missed (CE_{01}) in the P-Census. The implementation of this includes matching the P-Sample to the E-Sample enumerations in the matching universe in the sampled block cluster and then following up on the nonmatched E-Sample cases to determine if they are in fact a correct enumeration at the correct location. The nonmatched E-Sample cases determined to be correct may represent either a CE_{11} or a CE_{10}. Those nonmatched correct enumerations found to be nonmovers would reflect a missed enumeration in the P-Sample. Note that it is possible that the missed enumeration in the P-Sample could have been due to not obtaining an interview for a household. However, if the nonmatched correct enumerations are for movers, people who no longer live at that location, determining if they represent enumerations included in the P-Census is not possible in general. This is a consequence of the mover procedure used to identify matches in the P-Sample. Since the P-Sample includes inmovers, the ability to determine if the P-Sample would have enumerated movers at the location they were at on Person Interview Day would be required, but it is more likely than not that this other location is not in sample. Therefore, distinguishing between a CE_{11} or CE_{10} is possible for nonmovers but not for movers.

The estimate of EE includes all E-Sample nonmatches that are not identified as correct enumerations in the correct location. The E-Sample also includes the census enumerations with insufficient information and were not processed in the matching. The followup of the nonmatched E-Sample enumerations allows us to determine enumerations that should not have been included anywhere. However, for those enumerations at the wrong location, a determination of whether the enumeration was also included at their correct location is not possible without an additional search for duplicates. Those with an enumeration also at their correct location are considered EE_{10} or in other words a duplicate. Also, in general as discussed above, distinguishing whether enumerations at the wrong location would have been captured at their correct location by the P-enumeration is not possible since this is only implemented on a sample basis. Therefore, separate estimates of the terms comprising the estimate of EE is not a product of the processing for the DSE.

7. Overcoming Data Challenges

Estimating component errors presents data challenges. Enumerations that do not have sufficient information for matching and followup and enumerations that are not data-defined have been excluded from coverage measurement processing. In addition, the coverage measurement interview did not collect information regarding where a person should have been enumerated.

Currently methodology is being developed to overcome inflated estimates of erroneous enumerations. There are two types of data challenges for estimating erroneous enumerations. One challenge is to develop methods to process enumerations with insufficient information for matching and followup so that the matching team can determine whether they are correct or erroneous. These methods will provide the data for the estimation of the

$EE_{00\text{-}II}$ term in Table 1. The other challenge is to be able to determine when an enumeration is the only enumeration for a person, just in the wrong place. These methods are directed at estimating the WL_{11} and WL_{10} terms in Table 1.

Current plans include an investigation of broadening the criteria for sufficient information for matching and followup. A preliminary study of matching enumerations classified as insufficient information in the 2000 A.C.E. demonstrated that relaxing the criteria for the enumerations included in the processing had promise (Livermore Auer 2005). The results of the preliminary study have aided in the design of processing for data collected in the 2006 Census Test.

The coverage measurement questionnaire for the 2006 Census Test contains questions to collect information to determine the "correct place" that people in the E-sample should be enumerated. The processing will use the results of a search of all the census questionnaires in addition to data collected in interviews to determine where people should be enumerated. Unfortunately the test is limited to a site and a search of the whole U.S. cannot be done. A study with 2000 A.C.E. data is not possible because the A.C.E. questionnaire does not have the appropriate data to examine this.

Enumerations that cannot be resolved with new methodology will be treated as a missing data problem. Although coverage measurement surveys for previous censuses have had missing data and imputation methods to compensate, the design of the 2006 coverage measurement data collection asks more questions. The processing also will have the results of a search of all the enumerations for duplicates. The goal of the coverage measurement followup is to collect the information necessary to resolve uncertainty about whether an enumeration is correct or erroneous.

Currently there is much debate about whether estimates of the number of erroneous enumerations among those that are not data-defined are possible. This is the $EE_{00\text{-}NDD}$ term in Table 1. Determining whether every non-data-defined enumeration is correct or erroneous is not possible because of the lack of data. Methodology for estimating the number of correct or erroneous non-data-defined enumerations at an aggregate level has not been developed. The data appropriate to use in forming such an estimate is not clear. Varying degrees of data are available for the enumerations classified as non-data-defined. Sometimes the number of people in a household is known. Other times only that the housing unit is occupied is known. Sometimes it is unclear whether the housing unit is occupied. Of course, if the 2010 Census contains few enumerations in the category of non-data-defined, they will not be an issue.

A proposal in lieu of forming an estimate of the $EE_{00\text{-}NDD}$ term is to perform a sensitivity analysis. The sensitivity analysis would estimate the number of erroneous enumerations and omissions under a range of assumptions about the non-data-defined enumerations At one extreme is the assumption that all the non-data-defined enumerations are erroneous, or $EE_{00\text{-}NDD}$ =NDD. At the other extreme is the assumption that all of the non-data-defined are correct, or $EE_{00\text{-}NDD}$ =0. Another possible assumption is that the non-data-defined are erroneous at the same rate as the enumerations with insufficient information for matching.

The research with the coverage measurement data collected and processed in the 2006 Census Test will provide answers and insight. Also, an investigation of methods for estimating the number of erroneous enumerations by the cause, such as duplication, wrong location, and other reasons, will use these data. These results will be used in refining methods to use in the 2008 Dress Rehearsal and the 2010 Census.

References

Hogan, H. (2003) "The Accuracy and Coverage Evaluation: Theory and Design". *Survey Methodology*, 29, 2, 129-138.

Hogan, H. (1993) "The 1990 Post-Enumeration Survey: Operations and Results", *Journal of the American Statistical Association*, 88, 1047-1060.

Hogan, H. (1992) "The 1990 Post-Enumeration Survey: An Overview", *The American Statistician*, American Statistical Association, 261-269.

Kostanich, D., Whitford, D., and Bell, W. R. (2004) "Plans for Measuring Coverage of the 2010 Census". *2004 ASA Proceedings.* American Statistical Association. Alexandria, VA. CD-ROM, 1626-1635.

Livermore Auer, P. (2005) "Enumeration Status of Census 2000 Enumerations Deemed Insufficient Information for Matching and Followup". *2005 ASA Proceedings.* American Statistical Association. Alexandria, VA. CD-ROM, 2700-2707.

U.S. Census Bureau (2004) "Accuracy and Coverage Evaluation of Census 2000: Design and Methodology". DSSD/03-DM. Issued September 2004. U.S. Census Bureau, Washington, DC.

Table 1. Status of Population for Census, Eligibility for E-sample, Matching Universe by P-Census Status

Census	Eligible for E-sample	Matching Universe	P-Census: In	P-Census: Not In	
In	In	In	CE_{11}	CE_{10}	EE_{10}
			WL_{11}	WL_{10}	
		Not In	II_{01}	II_{00}	$EE_{00\text{-}II}$
	Not In		NDD_{01}	NDD_{00}	$EE_{00\text{-}NDD}$
Not In			OM_{01}	OM_{00}	