Chapter 6: A.C.E. Revision II Estimation

William R. Bell, Richard A. Griffin, Donna L. Kostanich,
Eric Schindler, Dawn E. Haines

The A.C.E. Revision II Dual System Estimate (DSE) methodology was developed with the following objectives in mind:

- Integration of the corrections for measurement errors so that we would not overcorrect for measurement errors identified by both the evaluations and the duplicate study.

- Separate estimation for both E- and P- Sample persons based on whether or not they linked to a census enumeration outside the search area.

- Flexibility in the post-stratification design, because the factors that affect correct enumeration (as measured by the E Sample) were not necessarily the same as those that affect coverage (as measured by the P Sample).

- Adjustment for correlation bias.

This chapter presents how this additional information was incorporated into the DSE for A.C.E. Revision II estimates. The reader is assumed to be familiar with the basic dual system model and how it was used to produce the March 2001 A.C.E. estimates. For a detailed description of this methodology see Haines (2001). The results of this methodology are presented in Davis (2001). Section 6.1 discusses the approach taken to dual system estimation and also describes the goals of the chapter. Section 6.2 discusses the estimation of the term accounting for persons in the Census who are not in the E Sample. The correct enumeration rate from the E Sample data is described in Section 6.3 while Section 6.4 addresses the estimation of the match rate from the P-Sample data. The census, E Sample, and P Sample data are combined in Section 6.5 to form a single DSE formula while Section 6.6 defines the post-stratification variables used for the A.C.E. Revision II Full and Revision samples. Section 6.7 discusses adjustment for correlation bias using demographic analysis sex ratios. Finally, Section 6.8 discusses synthetic estimation.

6.1 Dual System Estimation

The basic form of the dual system estimate (DSE) is:

$$DSE = (Cen' - II') \times \frac{CE}{E} \times \frac{P}{M} \qquad (1)$$

where Cen' = the census count, excludes late adds

II' = census records with "insufficient information" for matching, excludes late adds

LA = "late additions" to the census, i.e. records included too late for A.C.E. processing; primarily reinstated cases from the housing unit duplication study

CE = E Sample weighted estimate of census correct enumerations

E = E Sample weighted estimate of census total enumerations (with sufficient information for matching, excludes late adds)

P = P Sample weighted estimate of total persons

M = P Sample weighted estimate of matches to census persons.

DSEs were computed separately within post-strata. A post-stratum is a group of people defined by demographic and geographic characteristics who are assumed to have the same probabilities of inclusion in the census. Post-strata can also be defined to have equal probabilities of correct enumeration (given inclusion with sufficient information) in the census.

The DSE in (1) can be written as a function of the final census count, *Cen*, which includes Late Adds, and the following three rates:

$$DSE = Cen \times r_{DD} \times \frac{r_{CE}}{r_M} \qquad (2)$$

where $r_{DD} = (Cen' - II') / Cen$ is the census data-defined rate. The numerator excludes late adds but the denominator includes late adds.

$r_{CE} = CE / E$ is the E-Sample correct enumeration rate

$r_M = M / P$ is the P-Sample match rate.

The three rates can be interpreted as estimates of probabilities. Thus, within post-stratum,

- r_{DD} estimates the probability that a census person record has sufficient (and timely) information for inclusion in A.C.E. processing,

- r_{CE} estimates the probability that a sufficient information census person record is a correct enumeration, and

- r_M estimates the probability that a person in the P Sample is included in the census.

The interpretation of r_M may be less obvious than the other two; it is the sample-weighted proportion of P Sample persons who were also found in the census. The general independence assumption underlying DSE is that either the census or the A.C.E. inclusion probability are the same (both are not required). Assuming causal independence, the match rate r_M estimates the probability of census inclusion for the post-stratum.

Equation (2) also gives an interpretation of how the DSE constructs population estimates within a post-stratum.

- Multiply the census count (*Cen*) by the data-defined rate, r_{DD}, to estimate the number of census persons who are data-defined and, therefore, eligible for inclusion in the E Sample.
- Reduce this product by multiplying it by the estimated probability of correct enumeration, r_{CE}
- Increase this result by dividing it by the estimated probability of census inclusion, r_M

The primary tasks in developing DSEs at the post-stratum level are the estimation of the three rates involved. The estimate r_{DD} is straightforward because it's based on 100-percent census tabulations. See Section 6.2 for further detail. The development of the estimates r_{CE} and r_M is more challenging and is discussed in detail in Sections 6.3 and 6.4, respectively.

The different estimation tasks can be tackled one term at a time. Basically, the goal is to estimate the numerators and denominators of the terms r_{CE} and r_M. Since E, the estimated number of total census enumerations with sufficient information, is a simple, direct sample-weighted estimate, the challenges relate mostly to developing the estimates *CE*, *P*, and *M*. The estimation challenges for A.C.E. Revision II focus on accounting for: (*i*) information from the revised coding of the A.C.E. Revision Sample, (*ii*) information from the A.C.E. Revision II study of census duplicates, and (*iii*) different post-stratification schemes for the Full E- and P-Samples. The most difficult issue is (*ii*).

Before proceeding to a detailed discussion of the A.C.E. Revision II DSE components, consider the general nature of the estimator. While the basic DSE shown in equation (1) was applied in the 1990 PES (Hogan 1993), the March 2001 A.C.E. incorporated the modification called PES-C estimation. See Haines (2001) and Mule (2001) for details. This DSE had the general form:

$$DSE^{C} = (Cen' - II') \times \frac{CE}{E} \times \frac{P_{nm} + P_{im}}{M_{nm} + \frac{M_{om}}{P_{om}} P_{im}} \qquad (3)$$

where the following quantities are all P-Sample weighted estimates for the given post-stratum:

M_{nm} = estimate of matches to census persons for *nonmovers*
M_{om} = estimate of matches to census persons for *outmovers*
P_{nm} = estimate of total *nonmovers*
P_{om} = estimate of total *outmovers*
P_{im} = estimate of total *inmovers*.

Nonmovers, outmovers, and inmovers were defined with reference to their status in the period of time between Census Day (April 1, 2000) and the A.C.E. interview. Nonmovers were those who did not move during this period, outmovers were those persons who moved out of a sample block during this period, and inmovers are those who moved into a sample block during this period. Equation (3) estimated P-Sample matches (M) as the sum of estimated matches among nonmovers (M_{nm}) and estimated matches among movers. The number of mover matches was estimated as the product of an estimated number of movers (P_{im}) and an estimate of the mover match rate (M_{om} / P_{om}). Thus, P-Sample outmovers were used to estimate the mover match rate while P-Sample inmovers were used to estimate the number of movers. This approach implies that $P_{nm} + P_{im}$ should be used for the estimated total of P-Sample persons (P).

Equation (3) can be further expanded to include post-stratification subscripts. The Full E- and P-Sample post-strata are denoted by subscripts *i* and *j*, respectively. The census term was calculated for the cross-classification of *i* and *j* post-strata, denoted *ij*. The DSE formula, using version C for movers, with different post-strata for the E- and P- Samples is:

$$DSE^{C}_{ij} = Cen_{ij} \times r_{DD,ij} \times \frac{\left[\frac{CE_i}{E_i}\right]}{\left[\frac{M_{nm,j} + \left[\frac{M_{om,j}}{P_{om,j}}\right] P_{im,j}}{P_{nm,j} + P_{im,j}}\right]}$$

6.2 Estimation of r_{DD}

Recall the general form of the DSE in equation (2). This section discusses the estimation of the data-defined rate, or "DD-rate."

The DD-rate estimate (r_{DD}) is defined as $(Cen' - II') / Cen$ for a given ij post-stratum, where Cen', II', and Cen are defined from 100-percent census tabulations. At the post-stratum level, $Cen \times r_{DD}$ reduces to $Cen' - II'$. This suggests that an alternative to computing r_{DD} at a post-stratum level is to compute $Cen' - II'$ for all levels (e.g., demographic post-stratum groups within small geographic areas) for which estimates were to be computed, and then to adjust these quantities by the appropriate r_{CE} / r_M factors. This approach may be problematic, especially when applied to very small areas.

The problem with direct computation of $Cen' - II'$ for very small areas can be seen with the following hypothetical example. Suppose a particular small geographic area (e.g., a collection of blocks) has a high rate of imputation in the census, say 15.0 percent. Imputation rates will vary geographically, and high rates could result from a number of factors, such as difficulties getting access to housing units in secure communities or difficulties in hiring and retaining census enumerators in a particular area. In this hypothetical example, removing all imputations from the census count for the area by computing $Cen' - II'$ would reduce the census count by 15.0 percent. Subsequent multiplication by the r_{CE} / r_M factors and summing the resulting DSEs over post-strata may increase the population estimate from this base, but perhaps by no more than two or three percent (depending on the post-stratum composition of the area). The net synthetic DSE would thus be 12.0 or 13.0 percent lower than the census count. While this estimate could make sense if almost all the housing units for which persons were imputed were actually vacant (and this fact were not discovered in the census enumeration), it would not make sense if most of the units were occupied and the high rate of imputation resulted from other factors such as those suggested above. Calculating r_{DD} for post-strata and applying it synthetically avoids such problems in small area estimates, though perhaps incurring some error for larger areas for which the direct tabulation of $Cen' - II'$ would be sensible.

The data-defined rates, r_{DD} , are computed at the detailed post-stratum obtained as the intersection of the E- and P-Sample post-strata.

6.3 Estimation of r_{CE}

This section discusses the estimation of the correct enumeration rate, $r_{CE} = CE/E$. The Full E Sample post-strata are denoted by the subscript i. The Revision E Sample has post-strata denoted by i', where i' is based on collapsed post-strata i. This means that the Revision sample post-strata were obtained by collapsing the Full sample post-strata i. The correct enumeration rate is written:

$$r_{CE,i} = \frac{CE_i^{ND} \; f_{1,i'} + \widetilde{CE}_i^{D}}{E_i} \tag{4}$$

Note that the numerator term separates the E-Sample enumerations with a duplicate link to a census enumeration outside the A.C.E. search area, as identified in the duplicate study, from those enumerations without a link. The duplicate study used computer-based record linkage techniques to match the Full P- and E- Samples to census enumerations outside the search area. The census enumerations included those enumerations that were added too late to be included in the E Sample as well as those enumerations that were determined to be duplicates and were never included in the census.

The term CE_i^{ND} estimates the number of correct enumerations in the Full E Sample without duplicate links in post-stratum *i*. This term includes the probability of not being a duplicate, 1- p_t .

The component $\tilde{CE}_i^{D}$ represents the estimated number of correct enumerations in the Full E Sample with duplicate links in post-stratum *i* which are retained after unduplication. This term includes the probability of being a duplicate, p_t, as well as the conditional probability that an E Sample case is a correct enumeration given that it is a duplicate to another census enumeration outside the A.C.E. search area.

The total weighted number of persons in post-stratum *i* in the E Sample are denoted by E_i.

The double-sampling ratio factor $f_{1,i'}$ corrects for measurement error based on the Revision E Sample. It is a ratio of an estimate that uses the revised coding (indicated by *) to an estimate that uses the original coding. These adjustments, which are calculated for measurement error post-strata i′, are represented by:

$$f_{1,i'} \quad = \quad \frac{\dfrac{CE_{i'}^{ND*}}{E_{i'}^{ND}}}{\dfrac{CE_{i'}^{ND}}{E_{i'}^{ND}}} \quad = \quad \frac{CE_{i'}^{ND*}}{CE_{i'}^{ND}} \, .$$

P- and E- Sample cases with duplicate links were assigned a nonzero probability of being a duplicate, p_t . P- and E- Sample cases without duplicate links were assigned a p_t of zero. This probability is usually 0 or 1 for E- and P- Sample cases, but some duplicate links have a value in between, indicating less confidence that the link is representing the same person. These probabilities are also transferred to the E- and P- Revision Samples.

Although the duplicate study identified E- and P- Sample cases linking to census enumerations outside the A.C.E. search area, this study could not determine which component of the link was the correct one since there were no additional data collected to determine this. Assuming that the linked person does exist, the goal is to determine which of the two locations is the appropriate place to count the person. Since linked persons may be geographically close or far apart, this has implications for the degree of synthetic error. On the E-Sample side, this study does not identify whether the linked E-Sample case is the correct enumeration. Thus, it is necessary to estimate the following conditional probability:

z_t the probability that an E-Sample case is a correct enumeration given that it is a duplicate to another census enumeration outside the A.C.E. search area.

E-Sample Links

From the duplicate study, an estimate of correct census enumerations can be derived by considering the situation of the linked enumerations as well as assuming that each link represents one correct enumeration. This assumes of course that the link consists of true duplicates. These assumptions are used to estimate the contribution to correct enumerations from Full E-Sample cases with duplicate links, including those originally coded as correct as well as those originally coded as erroneous. This contribution to correct enumerations is given by the term: $C\tilde{E}^{D}{}_{i}$. To estimate this term, the E-Sample links are first classified according to the characteristic of the linked situation and the original coding of the E Sample. Attachment 1 summarizes this classification and the rules for assigning z_t's.

First, linked situations are identified where one component of the link is thought to be correct and the other incorrect. If a person in a housing unit links with a person in a group quarters, such as a college dormitory, the person in the housing unit is taken to be incorrect and assigned a z_t of zero. See "Linked Situation" 1. in Attachment 1. If a linked person 18 years of age or older is listed in only one of the households as a child of the reference person, this person is assumed to be incorrectly included with their parents and correctly included in the other household unless A.C.E. had already determined them to be an erroneous inclusion. An example of this might be a college student that was listed with their parents and also listed in an apartment off campus. This is represented by "Linked Situations" 2a. and 2b. in Attachment 1.

For other "Linked Situations" the choice of which person is correct is not clear. Consider links between whole households where all household members are duplicated. ("Linked Situation 3.) This includes families that might have moved some time around Census Day and were inadvertently included at both places or this might involve households with multiple residences with a helpful, but perhaps, uninformed proxy respondent. Another situation, "Linked Situation" 4., involves children ages 0 to 17, perhaps of divorced parents, that are linked between two different households. For these and all other situations, it is assumed that only half of these census enumerations with duplicate links are correct. To estimate the conditional probability, z_t , that the E-Sample person is the correct enumeration, controls cells are defined for "Linked Situations" 3., 4., and 5. as shown in Attachment 1 by:

- 3 Race/Hispanic Origin Domain
- Tenure

These resulting control cells are given in Attachment 2. Within each control cell the z_t 's are determined such that duplicate E-Sample cases originally coded correct or unresolved will weight up to one half the number of census duplicates identified including the erroneous enumerations. This was calculated as:

$$\hat{z}_t = \frac{0.5 \sum_t W_{.t}\ p_{.t}}{\sum_t W_{.t}\ p_{.t} \Pr(CE)}$$

The summations are over the links in a control cell regardless of the original E-Sample coding.

The components of Equation (4) are defined below.

$$\widetilde{CE}_i^{D} = \sum_{t \in i} W_{\pi,t}^{E}\ p_t\ z_t\ PRce_{\pi,t}$$

was the estimated number of correct enumerations with duplicate links in post-stratum *i* who were retained after unduplication.

z_t — the probability that an E-Sample case is a correct enumeration given that it is a duplicate to another census enumeration outside the A.C.E. search area.

$$CE_i^{ND} = \sum_{t \in i} W_{\pi,t}^{E} (1 - p_t)\ PRce_{\pi,t}$$

was the number of correct enumerations without duplicate links in post-stratum *i*, where the summation was taken over all enumerations in the A.C.E. E Sample in post-stratum *i*.

$W_{\pi,t}^{E}$ — was the production A.C.E. sampling weight for E-Sample person *t*.

p_t — was the probability that person *t* has a duplicate link outside the search area. This was usually 0 or 1, but could be between these two values for probability matches, where the accuracy of the link was uncertain.

$PRce_{\pi,t}$ — was the probability that person *t* was a correct enumeration in the original production coding. This was either 0 or 1 unless it was not possible to code the E-Sample case a correct or erroneous enumeration. In these cases a probability of correct enumeration was imputed.

$$f_{1,i'} = \frac{CE_{i'}^{ND*}}{CE_{i'}^{ND}} = \frac{\sum_{t \in i'} W_{RR,t}^{E} (1 - p_t)\ PRce_{R,t}}{\sum_{t \in i'} W_{R\pi,t}^{E} (1 - p_t)\ PRce_{\pi,t}}$$

$W^{E}_{RR,t}$ was the A.C.E. Revision Sample weight for person t to be used for Revision Sample coding.

$W^{E}_{R\pi,t}$ was the A.C.E. Revision Sample weight for person t to be used with production coding. These two weights could differ slightly depending on TES status and non-interview adjustment.

$PRce_{R,t}$ was the probability that person t was a correct enumeration in the A.C.E. Revision Sample coding.

$$E_i = \sum_{t \in i} W^{E}_{\pi,t}$$

was the total weighted number of persons in the E-Sample in post-stratum i.

6.4 Estimation of r_M

This section discusses the estimated match rate in Equation (2). While the E-Sample post-strata are indexed by i, the P-Sample post-strata are indexed by j. The match rate for post-stratum j is represented as:

$$r_{M,j} = \frac{M^{ND}_{nm,j} f_{2,j'} + \tilde{M}^{D}_{nm,j} + \left[\frac{M_{om,j} f_{3,j'}}{P_{om,j} f_{4,j'}}\right]\left(P_{im,j} f_{5,j'} + g\left(P^{D}_{nm,j} - \tilde{P}^{D}_{nm,j}\right)\right)}{P^{ND}_{nm,j} f_{6,j'} + \tilde{P}^{D}_{nm,j} + P_{im,j} f_{5,j'} + g\left(P^{D}_{nm,j} - \tilde{P}^{D}_{nm,j}\right)} \quad (5)$$

The residence status of P-sample movers was adjusted for coding error. The computer matching results were not used. Outmovers in the P sample were collected by a proxy interview, which made it difficult to obtain date of birth and age information. Since date of birth and age were important characteristics used in the computer matching, the movers were only adjusted for coding error.

Although the duplicate study identified E- and P- Sample cases linking to census enumerations outside the A.C.E. search area, this study could not determine which component of the link was the correct one since there were no additional data collected to determine this. Assuming that the linked person does exist, the goal is to determine which of the two locations is the appropriate place to count the person. Since linked persons may be geographically close or far apart, this has implications for the degree of synthetic error.

On the P-Sample side, this study does not identify whether the linked P-Sample case is a resident on Census Day. Thus, it is necessary to estimate the following conditional probability:

h_t is the probability that a P-Sample case is a resident on Census Day given that it links to a census enumeration outside the A.C.E. search area.

P-Sample Links

Unlike the E-Sample side, the duplicate study does **NOT** provide an estimate of the number of correct Census Day residents in the P Sample. In order to estimate h_t the probability that a P-Sample case is a resident on Census Day given that it links to a census enumeration outside the search area, it is necessary to borrow the resulting z_t's from the E-Sample links. Attachment 1 summarizes how the h_t's borrow information from the z_t's.

First, the P-Sample links to census enumerations outside the search area are identified for situations where it can be determined which component of the link is the correct residence. The "Linked Situations" and rules for assigning h_t's are the same as used for comparable types of E-Sample links. For example, consider a P-Sample person 18 years of age or older listed as a child of the reference person who links with a census enumeration in a household where they are not listed as a child, this P-Sample person would be assigned an h_t of zero regardless of how A.C.E. coded this person. Thus, it is assumed that this person should not have been included in the P Sample.

For the other "Linked Situations" 3., 4., and 5., there once again is no information to determine whether the P Sample had the person at the correct location or whether the census had them at the correct location. Additionally, there is no reasonable assumption about how many of these linked P-Sample persons should be at the correct location. To overcome this obstacle, it is assumed that the error in identifying correct residence is similar to the error in identifying correct enumeration for similar situations. Therefore, the h_t for P-Sample persons is set equal to the z_t determined for the E Sample for comparable linked situations as identified by the control cells in Attachment 2. The h_t's are then included in the weighted tallies, along with the p_t, to calculate the duplicate contribution to the Full P-Sample nonmovers and nonmover matches.

The terms in Equation (5) are defined below. Summation $t \in j$ denotes summation over A.C.E. P-Sample post-stratum j. Summation $t \in j'$ denotes summation over Revision Sample post-stratum j'. Summation is also indicated over nonmovers, outmovers, or inmovers, using either Production (π) or Revision (R) Sample coding.

$$M_{nm,j}^{ND} = \sum_{\substack{t \in j \\ t\ nonmover \\ production}} W_{\pi,t}^{P} \left(1 - p_t\right) PRes_{\pi,t}\ PRm_{\pi,t}$$

$W_{\pi,t}^{P}$	was the P-Sample production weight of person *t*.
p_t	was the probability that person *t* had a duplicate link outside the search area.
$PRm_{\pi,t}$	was the probability that person *t* was a match in the production coding.
$PRres_{\pi,t}$	was the probability that person *t* was a resident in the production coding.

$$f_{2,j'} = \frac{M_{nm,j'}^{ND*}}{M_{nm,j'}^{ND}} = \frac{\sum_{\substack{t \in j' \\ t\ nonmover \\ revision}} W_{RR,t}^{P} \left(1 - p_t\right) PRres_{R,t}\ PRm_{R,t}}{\sum_{\substack{t \in j' \\ t\ nonmover \\ production}} W_{R\pi,t}^{P} \left(1 - p_t\right) PRres_{\pi,t}\ PRm_{\pi,t}}$$

was the double sampling adjustment for nonmover matches.

$PRm_{R,t}$	was the probability that person *t* was a match in the Revision Sample coding.
$PRres_{R,t}$	was the probability that person *t* was a resident in the Revision Sample coding.
$W_{RR,t}^{P}$	was the A.C.E. Revision Sample weight for person *t* to be used for Revision Sample coding.
$W_{R\pi,t}^{P}$	was the A.C.E. Revision Sample weight for person *t* to be used with production coding. These two weights could differ slightly depending on TES status and the non-interview adjustment.

$$M_{om,j} = \sum_{\substack{t \in j \\ t\ outmover \\ production}} W_{\pi,t}^{P}\ PRres_{\pi,t}\ PRm_{\pi,t}$$

was the number of matched outmovers in the Full Sample in post-stratum *j*.

$$f_{3,j'} = \frac{M^*_{om,j'}}{M_{om,j'}} = \frac{\sum_{\substack{t \in j' \\ t\ outmover \\ revision}} W^P_{RR,t}\ PRres_{R,t}\ PRm_{R,t}}{\sum_{\substack{t \in j' \\ t\ outmover \\ production}} W^P_{R\pi,t}\ PRres_{\pi,t}\ PRm_{\pi,t}}$$

was the double sampling ratio for matched outmovers for post-stratum j'.

$$P_{om,j} = \sum_{\substack{t \in j \\ t\ outmover \\ production}} W^P_{\pi,t}\ PRres_{\pi,t}$$

was the number of outmovers in the Full Sample for post-stratum j.

$$f_{4,j'} = \frac{P^*_{om,j'}}{P_{om,j'}} = \frac{\sum_{\substack{t \in j' \\ t\ outmover \\ revision}} W^P_{RR,t}\ PRres_{R,t}}{\sum_{\substack{t \in j' \\ t\ outmover \\ production}} W^P_{R\pi,t}\ PRres_{\pi,t}}$$

was the double sampling ratio for outmovers for post-stratum j'.

$$P_{im,j} = \sum_{\substack{t \in j \\ t\ inmover \\ production}} W^P_{\pi,t}$$

was the number of inmovers in the Full Sample post-stratum j.

$$f_{5,j'} = \frac{P^*_{im,j'}}{P_{im,j'}} = \frac{\sum_{\substack{t \in j' \\ t\ inmover \\ revision}} W^P_{RR,t}\ PRinmover_{R,t}}{\sum_{\substack{t \in j' \\ t\,s\ inmover \\ production}} W^P_{R\pi,t}}$$

was the double sampling ratio for inmovers for post-stratum j'.

$PRinmover_{R,t}$ was the probability that person t in the Revision Sample was an inmover.

$g\left(P^{D}_{nm,j} - \tilde{P}^{D}_{nm,j}\right)$

The term g adjusts the number of inmovers for those Full P-Sample nonmovers who are determined to be nonresidents because of duplicate links. Some of these nonresidents are nonresidents because they are inmovers and should be added into the count of inmovers. The term $P^{D}_{nm,j} - \tilde{P}^{D}_{nm,j}$ is an estimate of nonresidents among nonmovers with duplicate links.

This term is multiplied by g, which is an estimate of the proportion of originally-coded nonmovers with duplicate links who are true nonresidents that have moved in since Census Day. The term g is estimated using the Revision Sample and both the original A.C.E. and the revision coding as follows:

$$g = \frac{P^{D}_{nm,im^*}}{P^{D}_{nm,nr^*}}$$

P^{D}_{nm,im^*} was an estimate of persons (using the Revision P Sample) with a duplicate link who were originally coded as a nonmover but the revision coding determined them to be inmovers, which are, of course, a subset of nonresidents.

P^{D}_{nm,nr^*} was an estimate of persons (using the Revision P Sample) with a duplicate link who were originally coded as a nonmover but the revision coding determined them to be nonresidents.

A couple of important assumptions are:

- If the revision coding determined a person was a nonresident, they really are a nonresident; i.e., revision-coded nonresidents are a subset of true nonresidents.
- The rate of inmovers for revision-coded nonresidents is the same as that for true nonresidents.

$$\tilde{M}^{D}_{nm,j} = \sum_{\substack{t \in j \\ t\ nonmover \\ production}} W^{P}_{\pi,t}\ p_t\ h_t\ PRm_{\pi,t}\ PRres_{\pi,t}$$

was the number of duplicate persons determined to have been Census Day residents who matched to the census.

$$P^{ND}_{nm,j} = \sum_{\substack{t \in j \\ t\ nonmover \\ production}} W^{P}_{\pi,t}\,(1 - p_t)\ PRres_{\pi,t}$$

was the number of non-movers without links outside the search area in post-stratum *j*.

$$f_{6,j'} = \frac{P^{ND*}_{nm,j'}}{P^{ND}_{nm,j'}} = \frac{\displaystyle\sum_{\substack{t \in j' \\ t\ nonmover \\ revision}} W^{P}_{RR,t}\,(1 - p_t)\ PRres_{R,t}}{\displaystyle\sum_{\substack{t \in j' \\ t\ nonmover \\ production}} W^{P}_{R\pi,t}\,(1 - p_t)\ PRres_{\pi,t}}$$

was the double sampling adjustment for nonmovers in post-stratum *j*′.

$$\tilde{P}^{D}_{nm,j} = \sum_{\substack{t \in j \\ t\ nonmover \\ production}} W^{P}_{\pi,t}\ p_t\ h_t\ PRres_{\pi,t}$$

was the estimated number of non-mover persons with duplicate links who were residents after unduplication.

6.5 The A.C.E. Revision II DSE Formula

The A.C.E. Revision II DSE formula, using version C for movers, separate E- and P- Sample post-strata, measurement error corrections from the E- and P- Revision Samples and Duplicate Study results is written:

$$DSE^{C}_{ij} = Cen_{ij} \times r_{DD,ij} \times \frac{\left[\dfrac{CE_i^{ND} f_{1,i'} + \widetilde{CE}_i^{D}}{E_i}\right]}{\left[\dfrac{M_{nm,j}^{ND} f_{2,j'} + \widetilde{M}_{nm,j}^{D} + \left[\dfrac{M_{om,j} f_{3,j'}}{P_{om,j} f_{4,j'}}\right]\left(P_{im,j} f_{5,j'} + g\left(P_{nm,j}^{D} - \widetilde{P}_{nm,j}^{D}\right)\right)}{P_{nm,j}^{ND} f_{6,j'} + \widetilde{P}_{nm,j}^{D} + P_{im,j} f_{5,j'} + g\left(P_{nm,j}^{D} - \widetilde{P}_{nm,j}^{D}\right)}\right]} \tag{6}$$

	Notation	
Terms	*CE*	Correct enumerations
	E	E-Sample total
	M	Matches
	P	P-Sample total
	f's	Adjusts for measurement error
	G	Adjusts nonmovers to movers due to duplication
Subscripts	*i,j*	Full E- and P- post-strata
	i',j'	Revision E- and P- measurement error correction post-strata
	nm, om, im	nonmover, outmover, inmover
Superscripts	*C*	DSE version C for movers
	ND	Not a duplicate to census enumeration outside search area
	D	Duplicate to census enumeration outside search area
	~	Includes probability adjustment for residency given duplication

PES-C was used because it was easier to count the inmovers than the out-movers in the P-Sample but it was easier to determine the match rate for outmovers. In some small post-strata, the number of inmovers was substantially larger than the number of outmovers. If there were only a few outmovers, the outmover match rate was subject to high sampling error. In these post-strata it was not considered appropriate to apply a suspect match rate to what could be a relatively large number of inmovers, so PES-A was used. PES-A uses only outmovers. PES-A was applied for post-strata with 9 or fewer P-Sample outmovers. For these post-strata, we made the assumption that some of the duplicate links determined not to have been residents were really outmovers.

The A.C.E. Revision II DSE formula that uses version A for movers with different post-strata for the E- and P- Samples is:

$$DSE_{ij}^{A} = Cen_{ij} \times r_{DD,ij} \times \frac{\dfrac{CE_i}{E_i}}{\left[\dfrac{M_{nm,j} + M_{om,j}}{P_{nm,j} + P_{om,j}}\right]}$$

The A.C.E. Revision II DSE formula, using version A for movers, separate E- and P- Sample post-strata, measurement error corrections from the E- and P- Revision Samples and Duplicate Study results is written:

$$DSE^{A}{}_{ij} = Cen_{ij} \times r_{DD,ij} \times \frac{\left[\dfrac{CE_i^{ND} f_{1,i'} + C\tilde{E}_i^{D}}{E_i}\right]}{\left[\dfrac{M_{nm,j}^{ND} f_{2,j'} + \tilde{M}_{nm,j}^{D} + M_{om,j} f_{3,j'} + g\left(M_{nm,j}^{D} - \tilde{M}_{nm,j}^{D}\right)}{P_{nm,j}^{ND} f_{6,j'} + \tilde{P}_{nm,j}^{D} + P_{om,j} f_{4,j'} + g\left(P_{nm,j}^{D} - \tilde{P}_{nm,j}^{D}\right)}\right]} \tag{7}$$

This version of the formula is used only when the sample size for outmovers in the Full P-Sample is strictly less than 10. This formula was used 93 times in the A.C.E. Revision II production process. The new terms introduced in this formula are defined as:

$$\tilde{M}_{nm,j}^{D} = \sum_{\substack{t \in j \\ t\ \text{nonmover} \\ \text{production}}} W_{\pi,t}^{P}\ p_t\ PRres_{\pi,t}\ PRm_{\pi,t}$$

was the number of matched P-Sample persons with duplicate links whether or not they were determined to have been residents by the unduplication process. Note that this formulation is different than that given for PES-C.

$$\tilde{P}_{nm,j}^{D} = \sum_{\substack{t \in j \\ t\ \text{nonmover} \\ \text{production}}} W_{\pi,t}^{P}\ p_t\ PRres_{\pi,t}$$

was the number of P-Sample persons with duplicate links whether or not they were determined to have been residents by the unduplication process. Note that this formulation is different than that given for PES-C.

6.6 A.C.E. Revision II Post-stratification Design

The Full E- and P- Samples with the original coding results that were used to produce the March 2001 estimates of census coverage provided the basis of the A.C.E. Revision II estimates. The

March 2001 A.C.E. estimates were determined to be unacceptable because of the presence of large amounts of measurement error. These Full Samples were comprised of over 700,000 sample persons each. Instead of one set of post-stratification variables, the A.C.E. Revision II estimates include separate post-strata for the Full E- and P- Samples, indicated by subscripts *i* and *j*, respectively.

6.6.1 Full P Sample

For the Full P Sample, the new post-strata were nearly identical to those used for the March 2001 A.C.E. estimates. The only difference was that the 0-17 age group was split into two groups, 0-9 and 10-17, which resulted in some collapsing differences. The Full P Sample, consisting of **480** post-strata, was based on the following characteristics (as opposed to the previous 416 post-strata):

- Race/Hispanic Origin Domain
- Tenure
- Size of Metropolitan Statistical Area
- Type of Census Enumeration Area
- Return Rate Indicator (Low vs. High)
- Region
- Age
- Sex

For the Full P Sample, the post-stratum groups either retained all 8 Age/Sex categories or were collapsed to 4 Age/Sex categories as shown below:

P - Sample Age/Sex Groupings

<table>
<tr><th rowspan="2">Age</th><th colspan="2">8 Groups</th><th colspan="2">4 Groups</th><th colspan="2">1 Group*</th></tr>
<tr><th>Male</th><th>Female</th><th>Male</th><th>Female</th><th>Male</th><th>Female</th></tr>
<tr><td>0 – 9</td><td colspan="2"></td><td colspan="2"></td><td colspan="2" rowspan="5"></td></tr>
<tr><td>10 – 17</td><td colspan="2"></td><td colspan="2"></td></tr>
<tr><td>18 – 29</td><td></td><td></td><td rowspan="3"></td><td rowspan="3"></td></tr>
<tr><td>30 – 49</td><td></td><td></td></tr>
<tr><td>50+</td><td></td><td></td></tr>
</table>

* The 1 Group is not used for the Full P post-strata (j), only the Revision P-Sample post-strata (j′).

Table 1 shows the 64 Full P-Sample post-stratum groups. The number in each cell represents the number of Age/Sex categories in each post-stratum group.

Table 1. Full P-Sample Post-Stratum Groups and Number of Age and Sex Groupings (j)

Race/Hispanic Origin Domain Number*	Tenure	MSA/TEA	High Return Rate NE	High Return Rate MW	High Return Rate S	High Return Rate W	Low Return Rate NE	Low Return Rate MW	Low Return Rate S	Low Return Rate W
Domain 7 Non-Hispanic White or "Some other race"	Owner	Large MSA MO/MB	8	8	8	8	8	4	8	4
		Medium MSA MO/MB	8	8	8	8	4	8	8	8
		Small MSA & Non-MSA MO/MB	8	8	8	8	4	8	8	8
		All Other TEAs	8	8	8	8	8	8	8	8
	Non-Owner	Large MSA MO/MB	8				8			
		Medium MSA MO/MB	8				8			
		Small MSA & Non-MSA MO/MB	8				8			
		All Other TEAs	8				8			
Domain 4 Non-Hispanic Black	Owner	Large MSA MO/MB Medium MSA MO/MB	8				8			
		Small MSA & Non-MSA MO/MB All Other TEAs	8				8			
	Non-Owner	Large MSA MO/MB Medium MSA MO/MB	8				8			
		Small MSA & Non-MSA MO/MB All Other TEAs	8				4			
Domain 3 Hispanic	Owner	Large MSA MO/MB Medium MSA MO/MB	8				8			
		Small MSA & Non-MSA MO/MB All Other TEAs	8				8			
	Non-Owner	Large MSA MO/MB Medium MSA MO/MB	8				8			
		Small MSA & Non-MSA MO/MB All Other TEAs	8				4			
Domain 5 Native Hawaiian or Pacific Islander		Owner	4							
		Non-Owner	4							
Domain 6 Non-Hispanic Asian		Owner	8							
		Non-Owner	8							
American Indian **or** **Alaska Native**	**Domain 1** (On Res.)	Owner	8							
		Non-Owner	8							
	Domain 2 (Off Res.)	Owner	8							
		Non-Owner	8							

6.6.2 Full E Sample

For the A.C.E. Revision II Full E Sample, the post-strata definitions have undergone major revisions. Some of the original post-stratification variables were omitted and additional variables were added. Logistic regression models identified several variables not included in the Full P-Sample post-stratification that were good indicators of correct enumeration. The Full E-Sample, consisting of **525** post-strata, was defined using the following characteristics:

- Proxy Status
- Race/Hispanic Origin Domain
- Tenure
- Household Relationship
- Household Size
- Type of Census Return (Mailback vs. Nonmailback)
- Date of Return (Early vs. Late)
- Age
- Sex

The new variables proxy status, household relationship and size, and type (mailback/non-mailback) and date (early/late) of census return are described generally below.

- Proxy Status - Non-proxy includes household members who lived there on Census Day and responded to the census (other than via an Enumerator Questionnaire).

- Household Relationship - The HHer/Nuclear relationship category includes persons in housing units consisting only of the householder with spouse or own children (l7 or younger). The "Other" relationship category consists of single-person households and persons in housing units with any other type of relationship, including unrelated persons.

- Household Size - Household size, or number of persons residing in the housing unit.

- Early/Late Mailback - Persons in mailback housing units with an earliest form processing date. On or before March 24 is early and after March 24 is late

- Early/Late Non-mailback - Persons in non-mailback housing units with an earliest form processing date: On or before June 1 is early and after June 1 is late

For the Full E Sample, the post-stratum groups either retained all 8 Age/Sex categories or were collapsed to 4, 2, or 1 Age/Sex groups, based on sample sizes, as shown below:

E - Sample Age/Sex Groupings

<table>
<tr><th rowspan="2">Age</th><th colspan="2">8 Groups</th><th colspan="2">4 Groups</th><th colspan="2">2 Groups</th><th colspan="2">1 Group</th></tr>
<tr><th>Male</th><th>Female</th><th>Male</th><th>Female</th><th>Male</th><th>Female</th><th>Male</th><th>Female</th></tr>
<tr><td>0 – 9</td><td colspan="2"></td><td colspan="2" rowspan="2"></td><td colspan="2" rowspan="3"></td><td colspan="2" rowspan="5"></td></tr>
<tr><td>10 – 17</td><td colspan="2"></td></tr>
<tr><td>18 – 29</td><td></td><td></td><td colspan="2"></td></tr>
<tr><td>30 – 49</td><td></td><td></td><td colspan="2"></td><td colspan="2" rowspan="2"></td></tr>
<tr><td>50+</td><td></td><td></td><td colspan="2"></td></tr>
</table>

Table 2 shows the 93 Full E-Sample post-stratum groups. The number in each cell represents the number of Age/Sex categories in each post-stratum group.

Table 2. Full E-Sample Post-Stratum Groups and Number of Age and Sex Groupings (i)

Proxy Status & Domain		Tenure	Relationship	HH Size	Early Mail-back	Late Mail-Back	Early Non-Mailback	Late Non-Mailback
Proxy: **Domain 7** Non-Hispanic White or "Some Other Race"					8			
Proxy: **Domain 4** Non-Hispanic Black					8			
Proxy: **Domain 3** Hispanic					8			
Proxy: **Domain 5** Native Hawaiian or Pacific Islander					1			
Proxy: **Domain 6** Non-Hispanic Asian					4			
Proxy: **Domain 1** American Indian or Alaska Native On Reservation					4			
Proxy: **Domain 2** American Indian or Alaska Native Off Reservation					1			
Non-Proxy: **Domain 7** Non-Hispanic White or "Some Other Race"		Owner	HHer/Nuclear	2-3	8	8	8	8
				4+	8	8	4	8
			Other	1	2	2	1	2
				2-3	8	8	2	4
				4+	8	8	4	8
		Non-Owner	HHer/Nuclear		8	8	8	8
			Other		8	8	8	8
Non-Proxy: **Domain 4** Non-Hispanic Black		Owner	HHer/Nuclear		4	4	2	4
			Other		8	8	4	8
		Non-Owner	HHer/Nuclear		8	8	8	8
			Other		8	8	8	8
Non-Proxy: **Domain 3** Hispanic		Owner	HHer/Nuclear		8	8	4	8
			Other		8	8	4	8
		Non-Owner	HHer/Nuclear		8	8	8	8
			Other		8	8	8	8
Non-Proxy: **Domain 5** Native Hawaiian or Pacific Islander		Owner & Non-Owner	HHer/Nuclear		2	2	2	2
			Other		2	2	1	2
Non-Proxy: **Domain 6** Non-Hispanic Asian		Owner & Non-Owner	HHer/Nuclear		8	8	4	4
			Other		4	4	2	4
Non-Proxy: American Indian or Alaska Native	**Domain 1** On Reservation	Owner & Non-Owner	HHer/Nuclear		8			
			Other		8			
	Domain 2 Off Reservation	Owner & Non-Owner	HHer/Nuclear		2	2	2	2
			Other		2	2	1	2

6.6.3 Revision P Sample

The Revision P Sample is a subsample of the Full P Sample and is comprised of over 70,000 sample persons. The Revision P Sample has been subjected to an additional field interview and/or rematching operation as part of the original A.C.E. evaluation program. In support of the A.C.E. Revision II program, the Revision P Sample has undergone extensive recoding using all available interview data and matching results. Missing data adjustments have also been applied to the Revision P Sample. This recoded data are used to correct for measurement error in the Full P Sample.

The measurement error correction post-stratum definitions (j′) depend on a person's mover status. Both Inmovers and Outmovers are subdivided into Owner and Non-Owner groups. For Nonmovers, the measurement error correction post-strata are: American Indians on Reservations (AIR) and, for the Non-AIR cases, a cross of Tenure (Owner versus Non-Owner) with eight Age and Sex categories. The Age/Sex collapsing pattern from the Full P Sample is retained when defining the measurement error correction post-strata. The Revision P-Sample Post-strata (j′) are defined as follows:

Mover Status & Domain	Tenure	**Age**	**8 Groups**		**4 Groups**		**1 Group**
			Male	**Female**	**Male**	**Female**	
Movers: **Domains 1 thru 7**	Owner						
	Non-Owner						
Non-Movers: **Domains 2 thru 7**	Owner	0 – 9					
		10 – 17					
		18 – 29					NA
		30 – 49					
		50+					
	Non-Owner	0 – 9					
		10 – 17					
		18 – 29					NA
		30 – 49					
		50+					
Non-Movers: **Domain 1** American Indian or Alaska Native On Reservation							

6.6.4 Revision E Sample

The Revision E Sample is a subsample of the Full E Sample and is comprised of over 70,000 sample persons. The Revision E Sample has been subjected to an additional field interview and/or rematching operation as part of the original A.C.E. evaluation program. In support of the A.C.E. Revision II program, the Revision E Sample has undergone extensive recoding using all available interview data and matching results. Missing data adjustments have also been applied to the Revision E Sample. This recoded data are used to correct for measurement error in the Full E Sample.

For the Revision E Sample, the measurement error correction post-strata are: Proxies, American Indians on Reservations (AIR) and, for the Non-Proxy/Non-AIR cases, a cross of a two-level Relationship variable with eight Age/Sex categories. Note that Household Size is collapsed out of the Household Relationship/Size variable. The Age/Sex collapsing pattern from the Full E Sample is retained when defining the measurement error correction post-strata. The Revision E Sample Post-strata (i′) are defined as follows:

<table>
<tr><th rowspan="2">Proxy Status
& Domain</th><th rowspan="2">Relationship</th><th rowspan="2">Age</th><th colspan="2">8 Groups</th><th rowspan="2">4 Groups</th><th rowspan="2">2 Groups</th><th rowspan="2">1 Group</th></tr>
<tr><th>Male</th><th>Female</th></tr>
<tr><td colspan="7"><u>Proxy:</u>
Domain 7 Non-Hispanic White or “Some Other Race”
Domain 4 Non-Hispanic Black
Domain 3 Hispanic
Domain 5 Native Hawaiian or Pacific Islander
Domain 6 Non-Hispanic Asian
Domain 1 American Indian or Alaska Native On Reservation
Domain 2 American Indian or Alaska Native Off Reservation</td><td></td></tr>
<tr><td rowspan="10"><u>Non-Proxy:</u>
Domain 7 Non-Hispanic White or “Some Other Race”
Domain 4 Non-Hispanic Black
Domain 3 Hispanic
Domain 5 Native Hawaiian or Pacific Islander
Domain 6 Non-Hispanic Asian
Domain 2 American Indian or Alaska Native Off Reservation</td><td rowspan="5">HHer/Nuclear</td><td>0 – 9</td><td colspan="2"></td><td rowspan="2"></td><td rowspan="3"></td><td rowspan="5">NA</td></tr>
<tr><td>10 – 17</td><td colspan="2"></td></tr>
<tr><td>18 – 29</td><td></td><td></td><td></td></tr>
<tr><td>30 – 49</td><td></td><td></td><td></td><td rowspan="2"></td></tr>
<tr><td>50+</td><td></td><td></td><td></td></tr>
<tr><td rowspan="5">Other</td><td>0 – 9</td><td colspan="2"></td><td rowspan="2"></td><td rowspan="3"></td><td rowspan="5"></td></tr>
<tr><td>10 – 17</td><td colspan="2"></td></tr>
<tr><td>18 – 29</td><td></td><td></td><td></td></tr>
<tr><td>30 – 49</td><td></td><td></td><td></td><td rowspan="2"></td></tr>
<tr><td>50+</td><td></td><td></td><td></td></tr>
<tr><td colspan="7"><u>Non-Proxy:</u>
Domain 1 American Indian or Alaska Native On Reservation</td><td></td></tr>
</table>

6.7 Adjustment for Correlation Bias using Demographic Analysis

The DSE estimates are adjusted to correct for correlation bias. 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 A.C.E. 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 A.C.E. Estimates of correlation bias are calculated using the “two-group model” and sex ratios from Demographic Analysis (DA). The sex ratio is defined as the number of males divided by the number of females. This model assumes no correlation bias for females or for males under 18 years of age; no correlation bias adjustment for Nonblack males aged 18-29; and that Black males have a relative correlation bias that is different than the relative correlation bias for Nonblack males. The correlation bias adjustment is also done by three age categories: 18-29, 30-49, and 50 and over. This model further assumes that relative correlation bias is constant over male post-strata within age groups. The Race/Hispanic Origin Domain variable is used to categorize Black and Nonblack.

The DA totals are adjusted to make them comparable with A.C.E. Race/Hispanic Origin Domains. Black Hispanics are subtracted from the DA total for Blacks and added to the DA total for Nonblacks. This is done because the A.C.E. assigns Black Hispanics to the Hispanic domain, not the Black domain. The second adjustment deletes the group quarters (GQ) people from the DA totals using Census 2000 data. The reason for making this adjustment is that the GQ population is not part of the A.C.E. universe. A final adjustment that could be made would be to remove the Remote Alaska population from the DA totals, since it too is not part of the A.C.E. universe. Since this population is small, the DA sex ratios would not be affected in any meaningful way. The resulting DA sex ratios for the three age groups by Black and Nonblack domain are shown in Attachment 3.

In general the correlation bias adjustment factor, c_κ , is defined for the three *k* age groups such that:

$$E\,[c_\kappa DSE^m{}_k] = \text{True male population for age group } k.$$

where:

$DSE^m{}_k$ is the sum of DSEs over male post-strata in age group *k*.

Since the purpose of this adjustment is to reflect persons missed in both the census and the A.C.E., the value of c_κ will not be allowed to be less than one.

<u>Correlation Bias Adjustment for Black and Nonblack Males 18 Years and Older</u>:

The correlation bias adjustment for Black and Nonblack males 18 years and older is done so that the A.C.E Revision II sex ratios will agree with the DA sex ratios for Blacks and Nonblacks.

This correlation bias adjustment is calculated as:

$$c_{R,\kappa} = \left(\frac{\sum_{ij \in k} DSE_{ij}^{Rf}}{\sum_{ij \in k} DSE_{ij}^{Rm}} \right) r_{DAR,\kappa}$$

where:

DSE_{ij}^{Rf} = DSE for race, R=Black or Nonblack, female post-strata ij.

DSE_{ij}^{Rm} = DSE for race, R=Black or Nonblack, male post-strata ij.

$r_{DAR,\kappa}$ = DA sex ratio for race, R=Black or Nonblack, for age group *k* as given in Attachment 3.

The sum over the *ij* post-strata includes only the intersection of those post-strata with age group *k*.

<u>DSEs adjusted for Correlation Bias:</u>

A correlation bias-adjusted DSE for a male 18+ post-stratum *ij* in the age-race group *k* is calculated as:

$$\widetilde{DSE}_{ij}^{m} = c_k \, DSE_{ij}^{m}$$

For all remaining post-strata, which includes female post-strata as well as post-strata for persons under 18 years of age, no correlation bias adjustment is done. Thus:

$$\widetilde{DSE}_{ij}^{f} = DSE_{ij}^{f}$$

The $\widetilde{DSE}_{ij}$'s are then used to form the synthetic estimates.

6.8 Synthetic Estimation

The coverage correction factors for detailed post-strata *ij* are calculated as:

$$C\tilde{C}F_{ij} = \frac{DS\tilde{E}_{ij}}{Cen_{ij}}$$

where:

$DS\tilde{E}_{ij}$'s are the correlation bias-adjusted DSEs for post-stratum ij.

Cen_{ij}'s are the census counts for post-stratum ij. Note that this Cen_{ij} includes late census adds.

A coverage correction factor was assigned to each census person excluding persons in group quarters or in Remote Alaska (effectively these persons have a coverage correction factor of 1.0). Recall that in dealing with duplicate links to group quarters persons, the person in the group quarter was treated as the correct enumeration or that this was their correct residence on Census Day. A synthetic estimate for any area or population subgroup *b* is given by:

$$\tilde{N}_b = \sum_{ij \in b} Cen_{b,ij}\, C\tilde{C}F_{ij}$$

Note that the coverage correction factor can be expressed as:

$$C\tilde{C}F_{ij} = \left(\frac{DD_{ij}}{Cen_{ij}}\right)\left(\frac{r_{ce,i}}{r_{m,j}}\right)c_k$$

where:

$r_{ce,i}$ is the correct enumeration rate component of the DSE, varying over *i* post-strata.

$r_{m,j}$ is the match rate component of the DSE, varying over *j* post-strata.

c_k is the correlation bias adjustment factor, varying over the Black and Nonblack groups and k age cells.

$\frac{DD_{ij}}{Cen_{ij}}$ is the data-defined rate, varying over the *ij* post-strata.

6.9 References

Davis, P. (2001), “Accuracy and Coverage Evaluation Survey: Dual System Estimation Results,” DSSD Census 2000 Procedures and Operations Memorandum Series #B-9*.

Haines, D. (2001), “Accuracy and Coverage Evaluation Survey: Computer Specifications for Person Dual System Estimation (U.S.) - Re-issue of Q-37,” DSSD Census 2000 Procedures and Operations Memorandum Series #Q-48.

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

Mule, T. (2001), “Accuracy and Coverage Evaluation Survey: Decomposition of Dual System Estimate Components,” DSSD Census 2000 Procedures and Operations Memorandum Series #B-8*

The “Linked Situations” and assignment of z_t’s and h_t’s occur in the order in which they are listed below.

Rules for Assigning z_t & h_t for Full P- and E- Sample Duplicate Links

	“Linked Situation” (E or P) ⇔ (Census)	**Original E Coding**	z_t	**Original P Coding**	h_t
1.	(Person in a housing unit) ⇔ (Person in a group quarters)	EE	0	NonRes	0
		CE/UE	0	Res/UE	0
2a.	(Person 18+, child of reference person) ⇔ (Person 18+, not child of reference person)	EE	0	NonRes	0
		CE/UE	0	Res/UE	0
2b.	(Person 18+, not child of reference person) ⇔ (Person 18+, child of reference person)	EE	0	NonRes	0
		CE/UE	1	Res/UE	1
3.	(All persons in a housing unit) ⇔ (All persons in another housing unit)	EE	0	NonRes	0
		CE/UE	$\hat{z}_1$	Res/UE	$\hat{z}_1$
4.	(Child 0-17) ⇔ (Child 0-17)	EE	0	NonRes	0
		CE/UE	$\hat{z}_2$	Res/UE	$\hat{z}_2$
5.	All Remaining Linked Situations	EE	0	NonRes	0
		CE/UE	$\hat{z}_3$	Res/UE	$\hat{z}_3$

EE is erroneous enumeration.
CE is correct enumeration.
UE is unresolved.
Res is resident on Census Day.
NonRes is not a resident on Census Day.

Attachment 2

Control Cells for Linked E Sample

Race/Hispanic Origin Domain	Tenure	"Linked Situation"	Control Cell
Domain 4 Non-Hispanic Black	Owner	3.	
		4.	
		5.	
	Non-Owner	3.	
		4.	
		5.	
Domain 3 Hispanic	Owner	3.	
		4.	
		5.	
	Non-Owner	3.	
		4.	
		5.	
Domain 7 Non-Hispanic White or "Some Other Race" **Domain 5** Native Hawaiian or Pacific Islander **Domain 6** Non-Hispanic Asian **Domain 1** American Indian or Alaska Native On Reservation **Domain 2** American Indian or Alaska Native Off Reservation	Owner	3.	
		4.	
		5.	
	Non-Owner	3.	
		4.	
		5.	

Correlation Bias Adjustment Groupings and Factors

Race/Hispanic Origin Domain	Age	DA Sex Ratios	Adjustment Factor
Black: **Domain 4** Non-Hispanic Black	18 - 29	0.90	1.08
	30 - 49	0.89	1.10
	50+	0.76	1.05
Nonblack: **Domain 3** Hispanic **Domain 7** Non-Hispanic White or "Some Other Race" **Domain 5** Native Hawaiian or Pacific Islander **Domain 6** Non-Hispanic Asian **Domain 1** American Indian or Alaska Native On Reservation **Domain 2** American Indian or Alaska Native Off Reservation	18 - 29	1.04	1.00*
	30 - 49	1.01	1.02
	50+	0.86	1.01

* This number set to 1.00 due to the inconsistency between DA and A.C.E. Revision II results.