Initial Assessment of the Census 2000 Accuracy and Coverage Evaluation Survey

Donna Kostanich, Rajendra Singh and Howard Hogan[1]
Bureau of the Census, Washington, DC 20233

Keywords: Dual System Estimation; Loss Function Analysis

1. Introduction

This paper summarizes the initial assessment of the Accuracy and Coverage Evaluation Survey (A.C.E). that was used to assist the Executive Steering Committee on A.C.E. Policy (ESCAP) in recommending whether to release data adjusted for census coverage error for purposes of redistricting. This is not intended as a comprehensive evaluation of Census 2000 or A.C.E. quality, but rather as a preliminary assessment for the redistricting decision that had to be made by March 2001.

Our goal was to do an initial assessment of the quality of the A.C.E. as compared to the quality of the Census. We were primarily concerned about the accuracy of congressional district data since the redistricting data files released in March of 2001 are used to form the new congressional districts. Since Census data have many other important uses, we also evaluated at the accuracy of state and county data. This initial assessment focused on the relative accuracy of population levels and population shares.

A loss function analysis approach was used, which was similar to how the 1990 undercount adjustments were evaluated (Mulry & Spencer 1993). Loss function analysis requires information on the errors present in the A.C.E. and the Census. The sampling errors for A.C.E. were known, however, at the time of this analysis, most of the non-sampling errors present in the A.C.E. were not known. Consequently, information on the pattern of non-sampling errors present in the 1990 Post-Enumeration Survey (PES) were used for this assessment of A.C.E. The loss function analysis considered the sensitivity of the non-sampling error assumptions. There were further limitations of the loss function analysis in that it did not directly reflect any errors due to the synthetic assumption that the net census coverage is uniform within a poststratum. A limited analysis on how synthetic error affects estimated losses was also studied.

Section 2 provides some background on the A.C.E. methodology. Section 3 reviews selected individual non-sampling error components of A.C.E. Section 4 synthesizes the error components into an indicator of overall relative accuracy of both the adjusted and the unadjusted census results.

2. A.C.E. Methodology

The A.C.E. Survey measures net coverage in the 2000 Census and can produce population estimates corrected for census coverage errors. The Census Bureau obtains a roster from the A.C.E. sample block clusters independently of the census. The independent roster (P Sample) is matched to the census to measure census omissions. A sample of census enumerations from the A.C.E block clusters (E Sample) is used to measure census correct enumerations. The results of matching and followup interviewing are used to determine which enumerations are the same person and which census enumerations are correct. After adjusting for missing data, Dual System Estimates(DSE) are calculated within population subgroups called post-strata. The DSE uses capture/recapture methods to estimate the true population. The DSE is the product of the number of census data-defined persons eligible and available for A.C.E. matching and the estimated proportion of correctly enumerated persons in the census (E-sample) divided by the estimated match rate from the A.C.E (P-Sample). Post-stratum DSE estimates are then used to determine coverage correction factors applied to all people counted in the census according to their specific post-stratum.

3. Select Non-sampling Error Components of A.C.E.

This section highlights some of the key non-sampling error components. We discuss specific issues related to each source of error and then contrast it to its 1990 PES analog. For the loss function analysis, the estimate of error due to these sources are mostly based on the 1990 PES evaluation data. (See B-1* for a more in depth discussion.)

3.1 Consistent Reporting of Census Day Residence

Proper application of the DSE model requires consistent reporting of Census Day residence between the P and E-samples. If a person who was sampled in the P-sample reports a different Census Day residence than he/she reported in the E-sample, then that person could be

[1]This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

considered both missed (based on the P-sample) and correctly enumerated (based on the E-sample), or conversely, both enumerated (based on the P-sample) and not correctly enumerated (based on the E-sample).

We expect consistency to be better than in 1990. First, the A.C.E. interviews were conducted much closer to Census Day than were the 1990 PES interviews, thus more accurate recall. In addition, the CAPI instrument forced the interviewers to ask all probes as to Census Day residence, again probably increasing consistency. On the other hand, the A.C.E. interview usually used proxy respondents for movers where the 1990 PES normally interviewed the mover household themselves. This could contribute to inconsistency; however, we have no direct data on this at this time.

3.2 Matching error

Matching error refers to assigning the incorrect code to a P-sample record. An error can be that a code of "matched" is assigned to a true nonmatch and vice-versa or that an unresolved code is assigned to a case that has sufficient information for matching. Matching errors can directly influence the final dual system estimates. Matching errors have both a random and a systematic component. The random component will be partially reflected in the overall variance estimates.

The matching error rate for 2000 is low with indications that it is substantially lower than the corresponding 1990 rate. In 1990, codes were entered into a computer system, but the actual matching and duplicate searches were done clerically using paper. The 1990 system attempted to match in-movers to their Census Day address. For the A.C.E., out-movers were matched within the sample area or its surrounding blocks. The 2000 matching system was significantly more automated, including computer assisted clerical matching, and all clerical matching operations were conducted at one location.

A matching quality assurance program provides information for assessing the matching operation, by measuring the level of error relative to that of our most experienced matching specialists. The results of the quality assurance process (see B-6*) show that we achieved a very high level of matching quality, with the majority of cases being computer matched.

3.3 A.C.E. Fabrications

Inclusion of fictitious people in the P Sample can create a bias in the DSE. Fictitious records have little chance of being matched which can erroneously increase the undercount estimates.

We have evaluated potential fictitious records in the A.C.E. by reviewing detailed quality assurance results that document the level of detected fabrications in the initial A.C.E. interview, as well as measures of residual fabrication. In addition we have the results of the Person Follow-up interviewing, which should have detected whole household P-sample fabrications not detected by the interviewing quality assurance program.

The evidence indicates that the quality assurance was successful in controlling A.C.E. fabrications. Because the A.C.E. interview was taken on the CAPI instrument, it was “time stamped” so that field staff could use automated reports to quickly detect interviewers who reported odd interviews, such as rapid multiple interviews, interviews at odd hours (such as late night interviews), and other similarly unbelievable interview results. The CAPI instrument allowed field management staff to tightly monitor the behavior of the A.C.E. interviewers.

In addition, we examined the data to look for information relating to clusters, because fabrication is often highly clustered. The matching analysts kept a detailed record of any unusual clusters and could request special questions during follow-up or send additional cases to follow-up interviewing if they questioned the integrity of one interviewer’s results. Analysts had the discretion to remove cases they believed to have been fabricated.

Based on the above discussion, we believe that the amount of fabrication in the P Sample is probably lower than it was for the 1990 PES.

3.4 Missing data

The A.C.E. maintained high quality interviewing and low levels of missing data. The level and pattern of missing data in the A.C.E. is comparable to that of the 1990 PES. (See B-7*)

Missing data can contribute to variance and, if the missing data models are poorly specified, can also contribute to bias and differential bias. A.C.E. has three types of missing data:

- Whole household noninterviews
- Unresolved match, residence, or enumeration status
- Missing demographic characteristics

This section focuses on the first two components of missing data. The third component of missing data may affect post-stratification and could result in correlation bias (see Section 3.7) or synthetic error (see Section 3.8). High levels of missing data, particularly for unresolved match, residence, or enumeration status, also tend to increase variance. We have not evaluated how this type of missing data by itself increases variance because this component is largely accounted for in our measure of sampling variance.

There have been two important changes for A.C.E. from the 1990 PES that might affect missing data rates. First, missing data could increase because of a different treatment of movers. In 1990 the PES only interviewed the current residents, whereas in A.C.E. we interviewed both the current residents and the Census Day residents. The Census Day residents include those who have moved out since Census Day and the day of the A.C.E. interview. A.C.E. interview rates were high. Among occupied housing units, the rates were 97.1 percent for Census Day and 98.8 percent for A.C.E. interview Day. This compares to 98.4 percent for interview day in the 1990 PES. Due to the high response, most of the changes due to noninterview adjustment factors applied were very small. This result helps to keep down variance of the survey weights.

For A.C.E, getting information about persons who had moved out was more difficult; however the matching of these individuals would be easier since the matching would be to census enumerations in the block cluster or a surrounding block. On the other hand for PES, getting information about only current residents was easier; however matching those residents who had moved in since Census Day was more difficult since their census enumeration would need to be geographically located at their Census Day address. This is an additional source of missing data. Second, the A.C.E. CAPI instrument kept the interviewer on the correct set of questions and allowed for tight managerial control.

The A.C.E. used a different missing data model for unresolved match and residence status. The 1990 model was based on hierarchical logistic regression, while the 2000 model used the far simpler "Imputation Cell Estimator." The input data and behavioral assumptions between the two models are similar but not identical.

3.5 Balancing error

Balancing error occurs when the set of correct enumerations records in the E-sample does not correspond to the set of records against which P-sample matching is allowed. An important type of balancing error occurs when the search area used to determine correct enumerations in the E-sample does not correspond to the search area used to determine matches in the P-sample. If the E-sample and the P-sample use different "search areas", this would imply different definitions of "correctly in the census," and the DSE model will not work.

The A.C.E. used a more complex balancing design than did the 1990 PES. The A.C.E. search area encompassed the first ring of blocks of housing units around the sample block. Not all cases were eligible for searching, coding, and matching in the surrounding ring; only whole household nonmatches and E-sample geocoding errors were eligible for surrounding block search. This search area is referred to as "Targeted Extended Search" or TES. The TES surrounding block search was also performed on a sample basis.

A major goal of extended search, whether targeted or not, is to reduce the variance of the estimators. Extended search can also reduce bias due to geocoding errors in the census and in the P-sample. If an A.C.E. address listing includes housing units outside the actual sample block, an attempt to match only to the sample block will result in nonmatches when the census units are also outside the block. This situation can lead to a high false measure of census omission. Extending the search to the surrounding blocks reduces this bias, essentially converting a first order matching bias to a second or third order sampling bias.

In addition, it is possible for the A.C.E. E-sample follow-up to incorrectly code a housing unit as inside a block when the unit is actually just outside the block. Without extended search, this discrepancy would result in a unit coded "correctly enumerated" that was actually a geocoding error. With extended search, the enumeration of the unit is correct whether coded to the actual block or a surrounding block. There is evidence that this type of coding sometimes occurred, as discussed in Section 3.6.

To assess the effect of TES, we compared correct enumeration rates and match rates for TES and non-TES cases. A review indicates an imbalance between P-sample matches to the surrounding block and E-sample enumerations coded as "correct in the surrounding block." In the absence of geocoding errors, these should be similar. This result raised concerns. However, it is consistent with the presence of a small amount of A.C.E. P and E-sample geocoding error. Similar results were encountered in 1990. An imbalance may be due to the geographic miscoding of E-sample cases discussed in Section 3.6.

3.6 Errors in Measuring Census Erroneous Enumerations

Census erroneous enumerations can occur if i) an individual had another residence where he or she should have been counted on Census Day, ii) a record is fictitious or duplicated, iii) an individual lived in a housing unit subject to geocoding error, or iv) the record had insufficient information for matching and follow-up.

Errors in measuring census erroneous enumerations can have a serious and direct impact on the A.C.E. For example, coding census fictitious cases as E-sample follow-up "noninterviews" leads to an incorrect estimate of the number of respondents correctly enumerated in the census. A tendency to "give the census the doubt" can result in people who move out before Census Day coded as correct enumerations. While the overlapping of the P and E-samples will lend considerable robustness to the A.C.E. estimates, both systematic and random errors can be expected to occur.

E-Sample cases are either coded during the initial matching operation or coded based on information gathered during A.C.E. follow-up. The identification of duplicates was closely monitored to assure that the duplicate search was done within the block cluster and in the surrounding blocks for TES clusters. The follow-up interview has been improved to instruct the interviewer to conduct sufficient searches for people to allow accurate coding of fictitious people.

For the A.C.E., we assessed errors in measuring census enumeration by analyzing the matching systems' quality assurance results, as well as by using information from A.C.E. follow-up. The in section 3.3 indicate that these processes were well controlled and that these errors were no worse than in 1990.

The one particular area of concern for A.C.E. is the level of correct coding of E-sample cases that were actually outside the search area. Preliminary results from an early A.C.E. evaluation indicate that a number of cases that were coded as "correctly enumerated" were in fact outside the search area. This means that the E-sample process accepted a number of records as correct when they were in fact erroneous. This would understate the gross census overcoverage rate and thus overstate the census net undercount.

3.7 Correlation bias

Correlation bias refers to error caused by individuals systematically more likely to be missed in both the census and the coverage measurement survey. This is a bias in the Dual System estimator, since dual system estimation assumes that the chance of being included in the P-sample is independent of the chance of being correctly included in the census. Correlation bias can occur from two sources. First, it can be caused by inherent and correlated heterogeneity within the post-strata. It can also arise when the event of being enumerated in the census changes the probability of being included in the A.C.E. Correlation bias will usually tend, therefore, to lead to an underestimate of the population. Dual system estimation will estimate some, but not all, of the people omitted from the census.

Correlation bias due to heterogeneity within post-strata can be reduced if either the census or A.C.E was more successful in including the "hard to count." The census paid advertising and outreach campaign, especially that targeted to ethnic minorities, could have reduced correlation bias in the 2000 DSE. However, the level of correlation bias in the A.C.E. might be larger than that in the 1990 PES because the A.C.E. was designed to reconstruct the Census Day household. This may increase correlation bias as it could result in missing individuals only tenuously connected to the household.

Analysis of correlation bias in the 2000 estimates was, as in 1990, based on the sex ratios from demographic analysis and is limited to only Black adult males and non-Black adult males. This analysis demonstrates the presence of correlation bias for adult Black males with levels comparable to that observed in 1990 and small levels of correlation bias for some adult non-Black males. We have no measure for correlation bias for children or females, nor any separate measure for Hispanics, Asians, or other separate "non-Black" groups. (See B-12*)

We could not estimate correlation bias for non-Black males 18-29 because demographic analysis sex ratios implied fewer males than measured by the A.C.E. This conclusion is important since the majority of the Hispanics, as well as of course other minority groups, are non-Black. A frequently expressed concern about the DSE methodology is the possibly large level of correlation bias for Hispanics.

3.8 Synthetic Assumptions

The A.C.E. coverage correction factors calculated for post-strata are applied to all census blocks within a post-stratum. This process is referred to as synthetic estimation, and a key assumption underlying this methodology is that the net census coverage is relatively uniform within post-stratum (i.e., the large area coverage rate is repeated in all contained small areas). Failure of this assumption leads to synthetic error. Synthetic estimation error differs from the other measurement errors discussed in this document because it is not directly related to the accuracy of the dual system estimates themselves, but rather to the distribution of the measured net undercount to local areas and demographic subgroups. Further, the heterogeneity exists in the unadjusted census as well.

Our assessment of the synthetic assumption looked at a bias correction for the loss function to see if the loss function results would be reversed. The findings showed minimal reversals.(See B-14*)

3.9 Other Measurement and Technical Errors

The coverage measurement process is subject to several other kinds of measurement errors, including technical ratio bias, contamination error, and inconsistent post-stratification. There was no evidence to believe that these errors were large enough to be of any concern (See B-1*).

Another concern has been the treatment in the DSE of those cases that were temporarily removed from the census because they were suspected of being duplicates. These records are referred to as "late census adds." These records were not included in the A.C.E. matching, processing, or follow-up processes. They were also excluded from the DSE. Some of these late adds would have been excluded from the DSE in any case (i.e., whole person imputations), some would have been erroneous enumerations and some would have been correct enumerations. It is possible that, had these records been included in the A.C.E. and the DSE, the estimated undercount would have differed.

If the ratio of matches to correct enumerations in the same for the excluded and included cases, the DSE expected value should be nearly the same. However, if the people referred to in the correct cases were either much more or less likely to have been included in the A.C.E., then excluding these cases from the A.C.E. would have changed the level of correlation bias and affected the A.C.E. We have no reason to believe this to be the case. Excluding these cases would have affected the sampling variances, especially if they were clustered. This effect, however, should be fully accounted for in the reported sampling error. If these late census adds included geographic clustering of erroneous enumerations, they would increase the geographic heterogeneity in the census net undercount. Geographic clustering in net undercount that is not correlated with the A.C.E. poststratification variables will not be corrected by the A.C.E. Even though there were late adds in 1990, the situation was entirely different and cannot be used to project the effect in 2000.

A related concern is that the level of whole person imputations in the census more than doubled from 1990 to 2000. The A.C.E. anticipated whole person imputations and was designed to exclude these cases from matching, follow-up and DSE. Some of these will be erroneous. Although they will not affect the DSE; they will decrease the net undercount rate and could lead to an overcount. Some of these imputation could be correct. Had these people been enumerated in the census and included in A.C.E., then some of them would have matched. Similar to the late adds scenario, if the ratio of matches to correct enumerations in the "imputed" cases is the same as the ratio matches to correct enumeration in the "non-imputed" cases, the DSE expected value should be the same. However, if people living in these units were either much more or less likely to have been included in the A.C.E., then imputing these cases (rather than enumerating them) would have changed the level of correlation bias and affected the A.C.E. Finally, the increased level of imputation would have affected the sampling variances, especially if they were clustered. This effect, however, should be fully accounted for in the reported sampling error.

Again, if incorrectly imputed cases were geographically clustered, they would increase the geographic heterogeneity in the census net undercount. Geographic clustering in net undercount that is not correlated with the A.C.E. poststratification variables will not be corrected by the A.C.E.

4. Synthesizing Quality

Loss functions provide a conceptual framework for assessing the accuracy of population estimates (Mulry and Spencer 1993) and are based on total error model. They assess the degree to which the adjusted and unadjusted data sets are closer to the target population for a specific grouping of areas (e.g., state, congressional district).

Let X and T denote vectors of population estimates and their target value, respectively. The loss function L(X,T) is a summary measure of the error in X as an estimate of T. X is more accurate than an alternative estimator Y if the expected value of L(X,T) is less than the expected value of L(Y,T). The general form of a loss function is given by:

$$L(X,T) = \sum_j w_j (X_j - T_j)^2 \qquad (4.1)$$

Where j denotes a geographic area such as state; X_j can be the census or adjusted census estimate of population level or population share: T_j is the target population level or share; and w_j is a weight which depends on the specific criteria. For estimating the loss of congressional district shares, we use the following form of the loss function:

$$L(X,T) = \sum_j Cen_j^2 \sum_i (X_{ij} - T_{ij})^2 \qquad (4.2)$$

Where X_{ij} is the population share of congressional district i within state j; T_{ij} is the corresponding target; and Cen_j is the census population for state j. Loss function (4.2) treats errors in all congressional districts the same regardless of the state a congressional district belongs to.

We prepared loss functions to determine the comparative accuracy of the adjusted and unadjusted data sets at the state and congressional district (CD) levels, to measure both numeric and distributive accuracy. See Census Bureau (2000) for important criteria in assessing accuracy. This Loss Function Analyses should not be considered determinative because we did not have complete information on A.C.E. biases.

To estimate the loss for the unadjusted and adjusted census data sets, one must properly account for several things. First is the estimated levels of undercount in the census as measured by the A.C.E. Second is the sampling variance in the A.C.E. Third is the level of bias present in the A.C.E. As we had no direct measures of the level of bias in the A.C.E. (except for ratio and correlation bias), we assumed the level measured in the 1990 PES. The analysis also took into account the variance of the estimated biases in 1990. A variety of models were run using different assumptions about the level of processing error in A.C.E. and the amount of correlation bias for adult men. (See B-13*)

Analysis shows that if A.C.E. processing errors are assumed at or near the level measured in 1990 and if there is little or no correlation bias, then either the unadjusted census is more accurate or the two are of nearly equal accuracy. If one assumes that the A.C.E. processing errors have been greatly reduced or if moderate or substantial correlation bias is present, then the A.C.E. adjusted results are more accurate, often by a large margin. Allowing for synthetic error does not reverse these findings.

If one assumes no reduction in processing error over 1990 as well as little or no correlation bias, the census is as accurate or more accurate than the adjusted A.C.E. for state levels, less accurate for state shares, and about as accurate for CD shares. This clearly demonstrates how sensitive the results are to the model assumptions. If the assumptions of similar patterns of errors do not hold even approximately, no direct conclusion can be drawn.

The loss functions run for counties with populations below 100,000 indicated that the unadjusted census was more accurate regardless of the level of correlation bias assumed. This caused some concern, since this was not the case for the 1990 census adjustment. One should remember, however, that counties below 100,000 are not the same or even representative of all areas of less than 100,000. However, the analysis found that the adjustment was more accurate when considered in terms of all counties for both numeric and distributive accuracy.

5. Conclusion

This preliminary analysis indicated that A.C.E. could potentially improve accuracy. These results are, however, limited to the extent that many of the non-sampling error parameters used in loss functions were based on the 1990 PES. This analysis should be repeated once the 2000 A.C.E. error parameters are available.

References (See B-1* for a more extensive list.)

_____, DSSD Census 2000 Procedures and Operations Memorandum Series B*, (B-1* to B-19*), all dated February 28, 2001.

Mulry, Mary H. and Bruce D. Spencer, "Accuracy of the 1990 Census and Undercount Adjustments," Journal of the American Statistical Association 88 (September 1993): 1080-1091.