Research Findings of the Accuracy and Coverage Evaluation and Census 2000 Accuracy

Howard Hogan*, Donna Kostanich, David Whitford and Rajendra Singh
U.S. Bureau of the Census, Washington, DC 20233-6200

Abstract: Evaluations of the Accuracy and Coverage Evaluation (A.C.E.) of the U. S. Census 2000 uncovered several problems leading the Census Bureau to decide to use unadjusted population counts instead of ones incorporating the A.C.E. results. This paper reviews the evaluation results, the design problems encountered, and our plans to produce revised estimates of Census 2000 coverage for purposes of potentially adjusting post-censal population estimates. This work on revising the A.C.E. estimates will lead to a coverage research agenda for the 2010 census.

Introduction

The U.S. Census Bureau conducted the Accuracy and Coverage Evaluation (A.C.E.) to measure the coverage of the population in Census 2000 and to allow for the possibility of correcting the census results for the measured undercount. The Census Bureau's Executive Steering Committee on A.C.E. Policy (ESCAP) recommended in March 2001 not to correct the 2000 Census counts for purposes of redistricting. The Secretary of Commerce concurred, and decided against adjustment. A major factor in the Census Bureau recommendation was the extreme difference between the A.C.E. and demographic analysis (DA). The March 2001 A.C.E. estimate of 3.3 million net undercount was much larger than the DA estimates then available. The Census Bureau position was that the possibility of an extreme error in the A.C.E. precluded the use of the A.C.E. for adjustment, and there was not sufficient time for further evaluation prior to the release of redistricting files. (See Hogan and Whitford, 2001)

The Census Bureau was still concerned about a potential differential undercount. Therefore, additional analysis were conducted over the next six months to determine if Census 2000 data should be adjusted for all uses other than redistricting. These evaluations indicated that there was a serious error in the A.C.E. In October 2001, the ESCAP II decided against adjusting for all other uses of Census 2000 data. The ESCAP II recommendations and rational, as well as all the supporting analysis are available on our web site. (www.census.gov)

Although unsuitable for census adjustment, the A.C.E. provides a wealth of information on the census process and may thus enable improvement in future censuses. We are now undertaking a revision of the A.C.E. estimates. These may also provide the basis of an adjustment of the base for post-censal estimates used by the Census Bureau and other Federal agencies throughout the decade.

This paper begins with a summary of key A.C.E. design decisions. It then summarizes the evaluation results and design problems encountered. It discusses our plans to produce revised estimates of Census 2000 coverage for purposes of potentially adjusting post-censal population estimates. This work on revising the A.C.E. estimates will lead to a coverage research agenda for the 2010 census. We will use results from this research to design a coverage measurement program in 2010. Additionally, we will explore using traditional coverage measurement techniques to improve census coverage quality while processing is underway.

Review of the A.C.E. Design

The A.C.E. was an implementation of the Dual-System Estimator within the Census context. It is a post-enumeration survey or coverage measurement survey. It was based on one number and two ratios within estimation cells. The number was the count of data-defined (that is, non-imputed) census enumerations. This number is tabulated directly from the census files.

The first ratio is the proportion of people who should have been included who were correctly included. This is the coverage ratio. It is estimated from a sample of people, known as the Population or P sample. It is measured by undertaking an initial interview, matching it to the census, following-up non-matches and doing a missing data adjustment. The initial interview was mainly conducted during June and July of 2000. The follow-up interview was conducted in October and November of 2000.

It is important that the P sample not be biased with respect to census coverage. An important bias is known as correlation bias, or failure of independence. In short, inclusion in the P sample can easily become correlated with inclusion in the Census, either through operational dependence or through response correlation.

The second ratio is the proportion of data-defined census enumerations that were correct. It is estimated from a sample of data-defined census enumerations, known as the enumeration or E sample. It is measured through duplicate search, follow-up interview and missing data adjustment. If a census enumeration matches a person listed in the initial (P-sample) interview, the information is used to code the enumeration as correct. Follow-up reinterview was also conducted in October and November of 2000.

The dual-system estimator is used to produce an estimate of the true population. It uses the product of the number of enumerations and the ratio of correct enumerations to total enumerations to estimate the number of people correctly counted in the census. This estimated number is divided by the coverage ratio (the ratio of correctly included enumerations to total people) to estimate the true number of total people.

To measure either the coverage ratio or the correctness ratio one needs a precise operational definition of a

correct enumeration. The Census Bureau has long used a definition that included the concept of correct location. Searching census records to see if a person had been enumerated is both expensive and time consuming, even with computer assistance. Accurately searching the entire country was deemed impossible. Searching many different locations was thought too expensive and difficult. Thus, the Census Bureau has restricted the search for a match to the area around where the person's usual residence was on April 1. A person counted else where is considered as not (correctly) enumerated. To balance the errors, that enumeration would be considered erroneous.

In Census 2000, the area around the April 1 usual residence included in the search areas was the sample block cluster and any block that touched the sample block cluster. A block cluster is usually a single census block but can be a group of contiguous blocks.

To see how this might work consider a simple example where there are fifteen blocks, each block a separate block cluster.

	A	B	C	

Assume that Block A fell into sample and assume the person lived in Block A on Census Day, and was interviewed by the A.C.E., P sample. If the person was enumerated by the Census in Block A, then we have one person, one enumeration and one match.

If the person should have been counted in Block A, but was counted in Block B, there would be one match, because we would look in the surrounding blocks and find the matching case. There would be no corresponding enumeration. However, if Block B fell into sample, that enumeration in that block would be considered correct as the person lived in a surrounding block.

If the person lived in Block A but was counted in both Block A and Block B, we would code the case as a match and directly identify the duplicate pair. So we would have one correct enumeration and one erroneous enumeration.

What happens if the person lived in Block A, but was only enumerated in Block C. In this case, searching would fail to find the person's corresponding census enumeration, because it was neither in the sample block cluster nor in a surrounding block. That is, it was outside the search area so we didn't look there. We would code the P-sample person as missed by the Census. If Block C fell into sample, we would code that enumeration as erroneous. The person was counted in a block outside the search area where he should have been counted.

Finally, if the person was counted in both Block A and Block C, he would be counted as correctly enumerated in Block A, and that enumeration would be coded as correct. The duplicate enumeration in Block C would be coded as erroneous.

Although somewhat complex, one can see that, assuming correct reporting and coding, there is a balance between P-sample misses and E-sample erroneous enumerations.

Making this determination can be quite difficult. Although in the schematic map above, "Block C" was close to the search area, in fact, the two enumerations could have been a hundred yards or a thousand miles apart. A person can be counted by the census in the wrong block for two distinct reasons. First, the person's housing unit may have been incorrectly coded by the census to the wrong block. This is known as a geocoding error. In this case the person may have always lived in Block A. Asking him when he moved or how long he has lived "here" is of no value. Field staff must visit the area and determine the correct location relative to the census geographic code.

The second instance is when the person has moved. Often the person moved after Census Day and was enumerated at his new home weeks or months later. Occasionally, the person mailed back a Census questionnaire and moved before Census Day. Other people have multiple residences between which they move. College students are an important example. Folks with vacation homes, some military, and "joint-custody" children may present similar problems.

In these cases the housing unit is correctly located by the census process. However, that unit was not the person's usual residence on April 1 according to Census rules. In these cases, we rely on respondent's answers to determine the correct location. In the case of a person counted in two non-adjacent blocks (e.g. A and C), we do not directly identify the duplicate. Instead, if Block A is selected, we code that enumeration as correct, based on responses obtained from that interview. If Block C is selected, we would code that enumeration as erroneous based on responses from that interview. We depend on sample size, weighting and estimation to balance the two.

This approach is known as unique location matching. As one can see, as applied here, it is quite dependent upon determining "Usual Residence on April 1" according to census rules.

Operationally, an enumeration could be classified as correct through two operation. An enumeration in the E sample was classified as correct if the enumeration was linked to a P-sample person listed as a Census Day residence during the initial A.C.E. interview (roughly May-July 2000). Non-matched E-sample cases were sent to a follow-up interview in October and November of 2000. At that time, field staff gathered information on the person's Census Day residence, and also on the physical location of the housing unit. Based on this information, clerks coded the E-sample case as correct or erroneous. On the P sample side, a matching case was accepted as "final." Many, but not all, non-matched cases were sent to follow-up, again to verify the information about the person's Census Day residence.

Finally, we treated some census enumerations as if they were not correctly enumerated, if they lacked sufficient

information for matching. This rule allowed us to implement precise matching. A similar rule was applied to the P sample. These rules increased independence. Before, when clerks or technicians made subjective decisions about which enumerations were "complete enough," they tended to selectively set aside non-matching cases and keep cases that matched. The decision of which cases were considered "in" was highly dependent on which cases matched. However, again, we increased the number of non-matching cases to be balanced by cases coded as erroneous.

An important design decision was how to treat movers. In both 1980 and 1990, interviewers asked about the people living at the sample address at the time of the PES interview. If current residents told us that they did not usually reside at the sample address on April 1, clerks looked at the reported Census Day location to determine whether they were counted or missed. However, regardless of that process, any enumeration of these non-Census Day residents in the sample block was coded as "erroneous."

For 2000, we changed how we treated movers. The key questions used to determine the match rate focused on the people who usually resided in the sample block on Census Day. This process requires accepting proxy interviews for most people who have moved out. In other words, we expected proxy respondents to know enough about their neighbors to apply census residence rules. However, while the neighbor may know that the family had a daughter of college age, the neighbor may not know whether the daughter lived at college (or was in the military or in jail, for that matter.) We do not yet understand how the interaction of "unique location matching," the treatment of movers, and proxy responses contributed to the error.

Finally, in 2000 the A.C.E. was restricted to the household population. All group-quarters populations, including students in dormitories, were excluded. The 1990 PES excluded only institutions and military group-quarters.

The A.C.E. estimation was done separately within 416 estimation cells. These cells were defined by demography (age, sex, race and Hispanic origin), tenure, type of census enumeration, and geography (census region and size of metropolitan area).

These estimation cells were developed from studying census evaluations since 1950, but especially the results of the 1990 Census. Once an estimated true population was obtained for a cell, the ratio was formed between this and the census count (including correct, erroneous and non-data defined enumerations.) This ratio, known as the coverage factor, was used to distribute the net undercount (overcount) to the local areas. (See Hogan, 2000 for a more detailed discussion)

Summary of Evaluation Results

The available evidence indicates that the estimate of the net undercount produced by the 2000 A.C.E. was much too high. The A.C.E. estimate of 3.3 million net undercount contrasts sharply with the current demographic analysis estimate of only 340 thousand. The ESCAP II report noted that the A.C.E. overstated the net undercount (and thus overstated the true population) by at least three million. (ESCAP II, p. i).

The evaluations addressed many aspects of the A.C.E. process and data quality. These are all available on the census web site (www.census.gov) under the ESCAP and ESCAP II headings. Here we will discuss the issues in terms of:

- Underestimation of census erroneous enumerations, due mainly to reporting error,
- Misestimation of census misses, also due mainly to reporting error,
- Errors due to geographic coding problems,
- Other issues.

Underestimation of census erroneous enumerations

Foremost, the A.C.E. failed to measure large numbers of census erroneous enumerations. This problem in the enumeration sample (E sample) seems largely due to errors in recording people's usual residences. This error alone was sufficient to call into question the survey results.

Two evaluation studies, Evaluation Follow-up (EFU) and Person Duplication, played a significant role in this finding. These are described here.

The EFU gave the Census Bureau the first indication that the A.C.E. was significantly underestimating erroneous enumerations. The evaluation consisted of an independent reinterview of 70,000 E-sample cases. i.e. census enumeration selected in the E sample and subsampled in the EFU. Census interviewers revisited the housing units to ask additional questions about residence. The field work was conducted in January and February 2001. (Krejsa, et. al., 2001, p. 3) The results were then coded by matching technicians.

This evaluation found an additional net 1,900,000 erroneously enumerated people. In addition, the EFU found cases that could not be resolved which represented about 4,600,000 people. (Krejsa, et. al., 2001, p.4) An unknown number of these unresolved cases, i.e. census enumerations, might have been erroneous.

Because of the "potentially significant implications" of these estimates, the Census Bureau undertook a very careful review of the EFU data and design. A "review sample" was chosen. The 17,000 "Review Cases" from the EFU were based on a probability sample selected from 70,000 E-sample cases that comprised the evaluation sample. This subsample of the subsample was recoded by our best coders. That is, more experienced coders were used to review the EFU questionnaires. No further field work was done. These coders determined a "best" enumeration status using both the original and reinterview results. Since this coding was intended as a check on the evaluation no missing data procedures were developed for cases that could not be resolved and the level of missing data was quite high. The revised estimate from this review was a net of 1,450,000 more erroneously enumerated people than found in the A.C.E. (Adams, et.

al., 2001, p. 1) That is, somewhat lower than the initial coding. However, the review took a conservative approach to coding difficult cases and concluded that there were over 15 million enumerations that could not be resolved or for which conflicting data had been collected.

A second evaluation study was a direct search for Census duplicates. For the first time the data capture of census names nationwide allowed these studies to apply computer matching to the entire nation. Duplication searches looked for all duplicates between the A.C.E. sample and the rest of the nation. It was known that this computer matching would miss many true duplicates (false non-matches), as well as identify a few false duplicate pairs. Since, the A.C.E. had done a complete (including clerical work) search for duplicates among data-defined enumerations within the sampled areas and their surrounding blocks, these data could be used to calibrate the efficiency of the national computer matching.

Mule (2001) matched census enumerations in A.C.E. block clusters to those across the entire country. Mule's matching consisted of a first stage nationwide match with strict rules followed by a within household match on any household with a match on the first stage. It was conservative in picking up census duplicates, because of his requirement for exact matching at the first stage. Within the A.C.E. block clusters, Mule found only 38 % of the duplicates that A.C.E. found, leading to the conclusion that the matching algorithm was underestimating duplicates in the census. Note that A.C.E. was not designed to estimate duplicates outside the search area and this itself was not a design flaw. The A.C.E. was however expected to determine which census enumerations were erroneous because they were reported at the wrong residence. Mule's study did not distinguish which of the duplicate pair was correct and which was erroneous, but one could easily speculate that half of these should be correct and half should be erroneous. This was not the situation for the census duplicates among the 17,000 cases coded in the EFU review sample. Of these duplicates, 70% were classified as correct by the evaluation review coders while only 16% were classified as erroneous. (Fay, 2002, p.23) Consequently, both A.C.E. and the evaluation follow-up failed to detect large numbers of erroneous enumerations.

By combining the EFU review results with the duplication study results, rough error in the A.C.E. estimates due to the measurement of erroneous enumeration, including duplication could be approximated. Fay produced lower bounds on the level of unmeasured erroneous enumerations of 2.9 million. (Fay, 2001, p.4). "Thus, the approximate range of potential overstatement of the net undercount was reduced to between 3 and 4 million persons." (ESCAP II, October 2001, p.11)

Misestimation of census misses

We discuss here the "misestimation" of census misses as the evidence indicates both significant numbers of false matches and false non-matches. We have not yet determined how these two "net-out" at either the aggregate or the post-strata level.

As noted above, the coverage ratio is calculated as the proportion of people who were correctly included in census. Thus, a person listed in the P sample is correctly enumerated only if he can be linked to a correct enumeration. A person listed in the Census with insufficient information or in an incorrect location (assuming that there was no other "enumeration") is treated as missed by the Census.

The evidence discussed in the previous section implies that many P sample cases are, in the production estimates, linked to erroneous enumerations. Remember, that many of the E-sample enumerations were coded as correct because they matched to a P-sample case. If we infer that the enumeration was erroneous, then that link cannot be valid. The extent of this problem has not been quantified.

However, other evidence indicates false non-matches as well. The EFU of P-sample cases provides some evidence. Nationwide computer search for both matched and unmatch cases, although preliminary, suggests problems in reporting usual residence similar to those that affected the E sample. Again the extent of this problem has not been quantified.

Errors due to geographic coding problems

One potential source of error we investigated was errors due to geographic coding, either errors not detected by the A.C.E. or introduced by the A.C.E. In a coverage measurement survey, the expected number of correct enumerations in the blocks surrounding the sample blocks should roughly equal the number of matches in surrounding blocks. The A.C.E. found about 3 million more matches in surrounding blocks than correct enumerations. (Navarro, et. al., 2001, p.9) Initially, we suspected that this balancing error might be a major problem in the A.C.E.

Immediately after the spring (ESCAP I) decision, the Census Bureau conducted field follow-up efforts to explore the balancing issue. Field staff checked the location of a sample of census housing units that had been coded as erroneous enumerations to determine if they were inside or outside of the sample block and surrounding ring of blocks. In addition, they checked units in the A.C.E. sample to see how often they were mistakenly included in the sample blocks but really existed in a block surrounding the sample block–called A.C.E. sample geocoding error. This effort determined that the major cause of the apparent balancing error derived from this latter type of geocoding error. Since in the A.C.E. matching, we searched the surrounding blocks for people, this type of error was insignificant; it had little or no effect on the net undercount estimates.

Other issues

Each of the other errors that can affect a dual-system estimator was also investigated, including correlation bias, missing data, contamination, and matching error as well as sampling variance. The level of error in each of these categories was approximately what we expected, and was

not a major basis of the decision.

One potential source of error that was troubling was synthetic estimation error. We became quite concerned that the sources of the additional erroneous enumerations were not even approximately uniform within the pre-defined post-strata. That is, past experience studying coverage patterns in censuses with significant net undercounts had lead us to define post-strata where with homogeneous chances of omissions. However, the nearly zero net Census 2000 coverage error was something quite new. Concerns were raised that our post-strata might be missing important variations in the net undercount, even if the omission chances were well modeled. We initially concluded that these concerns were not sufficient to reject adjustment. However, our research on synthetic estimation was based on the original estimates, i.e. estimates with a measured 3.3 million net national undercount. We have not yet reassessed the effect with revised, and much smaller, net national undercount.

Design Problems

The 1990 Post-Enumeration Survey (PES) seemed to have estimated the net undercount very well. Indeed, it produced an estimate of net undercount remarkably close to the current demographic analysis estimate for 1990, that is, the PES estimated 1.58 percent while demographic analysis sets the figure at 1.65 percent. (Robinson, 2001, p.4) All evidence indicates that the 2000 A.C.E. was operationally much better tested, was better controlled, had better Quality Assurance Procedures, and was better managed than the 1990 PES. What went wrong?

As noted above, our design has increasingly relied on a careful balancing of measured omissions and erroneous enumerations. This balancing has included not just complete misses and completely false or duplicate enumerations. It has also balanced those missed because they were not completely and correctly enumerated, using quite strict rules, against treating as erroneous all enumerations that do not meet these standards. This accumulation of, if you will, pseudo-erroneous inclusions may have kept us from seeing an increase in core erroneous enumerations, that is to say, people who were counted more than once as well as fictitious enumerations. In other words, the process of identifying and coding erroneous enumerations was increasingly seen through the need of balancing. We gave much thought to better measurement of missed people, much less thought to the measurement of true erroneous enumerations.

However, the system seemed to work and work well in 1990. So we must ask: "What changed?" Part of the problem may have been the changed treatment of movers. We evaluated this change mainly in terms of its effect on estimating the coverage ratio, that is balancing more accurate matching against possible losses of independence. However, since we use the initial interview to code the E sample as well, the change may have affected our estimated erroneous enumerations. Under the 1980 and 1990 approach, if the person says he is not living at the sample address on Census Day, we can code any enumeration there as erroneous whether we match the person at the address where he was living or not. Since these "in-movers" were not a prime focus of the A.C.E. interview, we may have suffered a loss in accuracy. We are investigating.

However, a review of A.C.E. procedures will only uncover part of the problem. After all, some of the errors we have discovered seem to be cases where the respondents consistently gave the "wrong" answer to the census, to the A.C.E., and to the evaluations. Something must be fundamentally wrong with our approach to discerning Census Day usual residences in both the Census and the evaluations.

"Unique Location Matching" does not necessarily require the coverage survey to adopt the Census residence rules. It only requires that each person be assigned one and only one location at which that person could be considered correctly interviewed/included/captured in the system. If the person was counted there, they were correctly enumerated. If they were counted elsewhere, that enumeration is erroneous. If the interviewers talked individually to each person (P and E samples), we could allow that person to choose any single location. However, we accepted both household and non-household proxy responses. Thus, we have tried to apply the Census concept of "usual residence" in the hope of getting more consistent responses. By consistency, we mean that the person (or those reporting for that person) will report the same usual residence no matter when and no matter where the interview is conducted. Evidence shows that the response to questions about usual residence depends upon where, that is, at which location, the interview is conducted.

The concept of "usual residence" was first adopted for the 1790 Census. It has a long history. As society has grown more complex, the rules that attempt to apply the concept have grown more complex. These rules may have become too separated from how many people view their lives. We have identified cases where the parents repeatedly and consistently report their college son or daughter as usually living at home, when in fact, the child lives at the university, according to census rules. Other evidence leads us to believe that both parents in a divorce (joint custody) situation reported their child as living with them. One wonders whether any amount of probing or clever questioning could force either parent to "reveal the truth," that is, the truth according to rules established by the Census. Clearly, usual residence was a large part of the problem.

A lesser, but still important part, was the failure to consistently apply the concept of search area. Again, the A.C.E. design required each housing unit to be assigned to a search area consistently, whether it was enumerated or missed in the Census and whether it was listed or missed in the A.C.E. However, in cases of "bad" maps or confusing geographic features, a single unit would (on a probability or expected value basis) have been coded differently in each case. Thus, some units would (again on an expected value basis) have been counted as both correctly included (E sample) and erroneously missed (P sample) in the Census. This contributed to the errors in the A.C.E..

One source of potential A.C.E. error may not have been major, but is worth mentioning. The 2000 A.C.E. excluded all group quarters. The 1990 PES included non-military non-institutional group quarters, including for example, small group homes as well as college dormitories. The change in design from "Who is living here now and where did they live?" to "Tell me who lived here on April 1?" seemed to make independent interviews at groups quarters difficult, if not impossible. However, this change had at least two costs. First, we only measured part of the net undercount. Reconciling the A.C.E. results with the DA results was more difficult. Also, the classification of small home-like group quarters was sometimes inconsistent between the A.C.E. and the Census. This inconsistency caused disproportionate operational difficulties.

Research Plans for Revised 2000 Estimates

Challenges

As mentioned above, as part of the October decision, Fay (2001) produced a lower bound on the level of unmeasured erroneous enumerations of 2.9 million. In addition, Thompson et. al., (2001) used this lower bound to produce a "Revised Early Approximation" of undercount for three race/ethnic groups to be illustrative of net undercount and possible coverage differences. The same methodology and data were later used to expand the calculations to seven race/ethnic groups. (Mule, 2002) Although these approximations are useful in putting the undercount into perspective, they are insufficient for the purpose of correcting the post-censal estimates base.

The "Revised Early Approximation" suffers from a number of limitations and was only intended to be illustrative of possible coverage differences. These approximations only partially corrected for measurement error in the A.C.E. estimate of census erroneous enumerations. Furthermore, these corrections were based on an extremely small subsample of about 17,000 persons, whereas the original A.C.E. results were from about 713,000 persons. Consequently, we could not get approximations for any more than the seven population groups.

The Census Bureau is in the process of developing revised estimates of the net undercount. Our plan calls for using methodology similar to that used by Thompson et. al., (2001) and Mule (2002). The revised A.C.E. estimates will make use of the original A.C.E. data, a subsample of the A.C.E. that has been corrected for measurement error, and a study that identifies and estimates duplicate census enumerations.

We have identified the following five challenges that we would need to overcome in order to succeed in developing revised A.C.E. estimates:

- Improved estimates of erroneous census enumerations.
- Improved estimates of census omissions.
- Better models for missing data.
- Enhanced poststratification.
- Adjustments for correlation bias.

It is critical to obtain better estimates of the level and distribution of census erroneous enumerations, and to do so for more detailed population groups. Fay's early revisions were based on a lower bound for erroneous enumerations and only for several population groups. To obtain better estimates of omissions, measurement error in the P sample, particularly residency status, needs to be addressed.

Many of the suspected E-sample duplicates were in fact matched to a P-sample case. The level of unresolved E-sample cases in the early revisions is unacceptably too high and there will be new challenges in dealing with missing data if the P-sample component is corrected. If a P-sample household was erroneously included, this will create a new type of noninterview adjustment not previously dealt with.

The causes and factors related to erroneous census enumerations are likely to be different than those related to census omissions. This implies considering separate post-stratification variables for the P and E samples to be able to capture those factors related to overcounts and those related to undercounts. In the presence of overcounts, correlation bias creates some new issues. In the past, we have assumed that ignoring correlation bias will give us a conservative estimate of the undercount; i.e., the adjustment is in the right direction, but just not enough. With overcounts this assumption may no longer be valid. It's now possible that a correction might not even be in the right direction. A further challenge will be dealing with variance increases since much of the information needed to correct the errors in the A.C.E. is based on relatively small samples.

We cannot go back and revisit households to obtain any additional information to correct the A.C.E. data. We will have to make do with the data collected in the original A.C.E. person interview, and data collected in the evaluation follow-up. Our objective is to obtain better estimates of census erroneous inclusions, and census omissions in order to calculate dual-system estimators for finer population subgroups. In general, our strategy will be to start with the original A.C.E. data and apply adjustments to the components in the dual-system estimate. The purpose of each adjustment is to correct for measurement error in that particular component, i.e., the correct enumeration rate. Adjustments will be based on a subsample that has been corrected for some measurement error using information from A.C.E. evaluations. Other adjustments to reflect errors in enumeration or residency status not detected by A.C.E. evaluations will be based on a study of census duplications, and a study of P-sample cases that match to a census enumeration outside the search area. There will likely be different post-strata for the original E & P samples, and fewer post-strata will be used when applying the adjustments.

Current Strategy

We began by analyzing the entire evaluation sample, which consists of about 70,000 E-sample persons. Using the results of the original A.C.E. interview, and the evaluation reinterview, a more accurate classification of whether a case is correct or erroneous can be made. This is comparable to the "best" enumeration status that was previously obtained for the 17,000 cases in the subsample of the EFU subsample. Furthermore, most of the matching can be done using a computer algorithm and only some of the cases will have to be clerically coded by our best analysts. This should provide substantial reductions in variance although there is a risk of some misclassification. A similar type of analysis was also conducted for P-sample cases that were part of the evaluation follow-up reinterview. On the P-side "best" codes will be determined for both residency and mover status. Some geocoding errors will also be corrected as part of this process. The sample with these corrected measurement codes is referred to as the A.C.E. Revision Sample.

Preliminary analysis of the A.C.E. revision data indicates that the proportion of unresolved cases will decrease considerably. Missing data models are being developed for remaining unresolved cases, and for conflicting cases where two apparently good interviews result in different classifications. Household noninterview adjustments are being developed for P-sample cases that get changed to nonresidents on Census Day.

Another major effort is underway to develop better methods to identify and estimate duplicate census enumerations. This will involve creating methods or models to increase the efficiency of the matching algorithm. The previous duplicate study, which used exact matching at the first stage, only identified 38% of the in-scope duplicates within a block cluster. Statistical matching algorithms will be researched to improve this efficiency. Additional models will be explored in an attempt to estimate the probability of a duplicate. We will also consider computer matching the P sample against the entire Census to obtain information about how well residency status is being measured.

We will not be able to correct for all measurement error in the A.C.E. Revision Sample. In particular, we will need to estimate census duplicates reflected in the A.C.E. Revision, and integrate these into the revised DSEs. Current plans assume that this will follow methodology similar to that used for the "Revised Early Approximation." A double sampling ratio adjustment will make use of the original A.C.E. sample to help reduce variance from the smaller A.C.E. revision. Post-strata will be re-examined. For the original A.C.E. data, separate E & P post-strata are being looked at, since there are different factors related to erroneous inclusions and omissions. Using different age categories is also under consideration. Much of the previous work on developing post-strata focused only on the census omissions, and by default the same post-strata were applied to the erroneous inclusions.

Post-strata will have to be aggregated to higher levels when estimating components based on the A.C.E. Revision Sample. Other possible adjustments may be made for P-sample people who link with census enumerations outside the search area, and to correct for correlation bias. Ignoring correlation bias will cause us to underestimate the population, which might have serious consequences when the true undercount is very near and sometimes below zero. It may also cause us to underestimate the differential undercount.

Until revised A.C.E. estimates are produced and assessed, we do not know if these will be able to provide improvements to population estimates. Variance estimates and other key quality indicators will provide a basis for whether the intercensal estimates are corrected. Regardless, the result of this work is critical to the planning of a better, more accurate, census, and the appropriate measurement of coverage in 2010.

Designing a 2010 Coverage Evaluation Program

In turning our attention to the planning of Census 2010, we would like to set forth the goals, questions and design principles that should guide our early thinking for the coverage evaluation studies. We will briefly discuss some ideas of how we can use matching, reinterview, and other coverage measurement techniques as part of the census itself in order to prevent coverage errors from occurring. We then focus on the independent coverage evaluation.

In Census 2000 we have already seen coverage measurement techniques being used in real time to correct the census. The efforts to identify and remove housing unit duplication relied on years of research in computer matching, research first developed for coverage evaluation. Considering that current estimates show nearly zero net undercount, one might well speculate upon the situation if these duplicates had not been removed, and the census produced a large net national overcount.

Secondly, real time targeted coverage surveys could be used as the eyes and ears of Census management. With our increased ability to turn these surveys around very quickly, we could run a coverage measurement survey (on a sample basis) to point to coverage errors (over as well as undercounts) in sufficient times to take corrective actions. Of course, considerable additional work needs to be done before we can conduct, even limited, real time coverage studies.

In 2000, we were reluctant to feed the results of the A.C.E. back to Census management to take corrective action for fear of either compromising statistical independence or giving the appearance of loosing organizational independence. Special targeted surveys, independent of the overall coverage evaluation, would not suffer this disability. We would also draw on early demographic-based benchmarks that provide indicators of the consistency, and completeness of the emerging census results.

This work should focus on very quick turn-around, especially during the census testing and dress-rehearsal cycle. It should be able to provide information about, for

example, the coverage patterns of the '06 test in time to influence the design of the '08 dress rehearsal test. It should provide data on the dress rehearsal in time to identify any major issues that must be corrected.

The Census Bureau is already looking at both these approaches. As plans develop, there will be serious decisions to be made with respect to allocation of resources between real time improvement, and completely independent evaluation. However, these are the right questions.

Recent coverage evaluations have focused on the net undercount for relatively large geographic and demographic groups. This information will continue to be of interest to the data user, and so must remain a major goal. The coverage evaluation should focus as well on increasing our understanding of the census process. Thus, it should provide more information to the designers of the census. It should be able to answer the questions that data users and census designers most frequently ask. We list some of these questions below.

The coverage evaluation should be designed to answer questions such as:

- How many people were missed?
- How many people were counted twice?
- How many people were counted in the wrong location (block, tract, county, state)?
- How many people were missed because their dwelling unit was missed?
- How many people were missed in enumerated dwelling units.
- How many people were duplicated in correctly enumerated dwelling units ("apartment mix-up")? What was the true population of those units?
- Did, on average, the household population imputation reflect the number of people in the housing unit?
- How many housing units were missed? How many group quarters? How many such units were included twice?
- What census processes caused, facilitated, or failed to detect these coverage errors? For example, how do these errors relate to enumeration method, LUCA, Be-Counted, PSA?
- How do these errors relate to social, economic or housing conditions?

In addition, the coverage evaluation must continue to provide information on the net undercount for:

States
Large Cities
Urban/Rural
Major race and ethnic groups
Age and sex

While we will need to refine and develop our design as the decade progresses, we should begin with the following design assumptions:

- We should include both housing unit, and the non-institutional group quarters.
- We should conduct careful address (dwelling unit) matching.
- Liberal rostering: We should include everyone with any reasonable attachment to the dwelling unit.
- We should interview the people who are present at the time of the coverage evaluation interview, and determine the place or places where they resided or might have been enumerated on Census Day (that is, use the 1990 approach).
- We should determine all locations where the person could have been enumerated and match there ("any location matching").
- We should conduct a full extended search, perhaps expanding the search area to two or more rings everywhere.
- We should continue to use field automation (including follow-up), computer matching and computer assisted clerical matching.

What early research is needed?

- We need more cognitive research on the concepts of usual residence, and this research must be integrated into the main census design. We did insightful research in the 1990's, but its effect on the basic census concepts was marginal at best.
- We need to reconsider what questions need to be "QA'ed" as well as which cases to send to follow-up. Specifically, we may want to send to follow-up a complete sub-sample, including matched cases.
- We need additional research on how to conduct the follow-up interview, including what questions to ask and how to automate.
- We need further research on measuring erroneous enumerations, including wider computer searching.
- We need to reconsider the treatment of census cases with incomplete information. Specifically, we want to research methods that utilize address matching to minimize our dependence on balancing "insufficient information" cases with "non-matches."
- We need to research missing data methods, especially methods that complement the "any-location" matching design.
- Is our approach to post-stratification correct? For example, do we need different post-strata to measure omissions and erroneous enumerations?

These goals imply a design quite different from the one used in the last three censuses. At the very least, the dual-system estimate will have to be modified to be less reliant on balancing. Household and housing unit concepts need to be explicitly worked into the design. However, freed from the requirement to produce early adjusted census figures, we can now more freely and fully research these approaches.

Additionally and quite importantly, both the Census itself, and the coverage measurement need to conduct extensive research on the public's concept of residence and residence rules, as well as research on methods to

facilitate recall. Indeed, we need to focus as much attention on measuring erroneous enumerations as on measuring omissions. This research must include extensive cognitive testing.

Conclusion

Let us add one further note. Acting Director William Barron has publicly stated that the Census Bureau has no business being in the middle of a process as intensely political as Congressional Redistricting.

> With respect to future uses of data for what I'll call political purposes such as redistricting, I think it's unlikely that the A.C.E. process could be made to work in a timeframe consistent with current law, and that is both a timing issue and a substantive technical issue. …. We've now had two decades where this has been tried, and it's helped to create an environment where census employees are forced to work in an atmosphere that I think is not conducive to good professional work, and I would like to see that change for the future and – a census is tough enough without the kind of political maelstrom that seems to associate itself with this particular issue. And given that it doesn't seem possible, I wish it would just stop. That's my view. [ESCAP II PRESS CONFERENCE, Suitland, Maryland, October 17 2001, transcript dated October 24.]

Although this is not a technical point, we think this point must be included in any discussion about what went wrong in 2000. Our focus on census adjustment has been costly in many ways. One of the costs is that it has led us away from important issues in evaluating and understanding census coverage.

References

ESCAP II, *"Executive Steering Committee for Accuracy and Coverage Evaluation Policy on Adjustment for Non-Redistricting Uses,"* October 17, 2001

Adams, T. and Krejsa, E. *"ESCAP II: Results of the Person Follow-up and Evaluation Follow-up Forms Review,"* Report 24, October 15, 2001.Fay, R., *"ESCAP II: Evidence of Additional Erroneous Enumerations from the Person Duplication Study,"* Report 9, October 26, 2001.

Fay, R., *"ESCAP II: Evidence of Additional Erroneous Enumerations from the Person Duplication Study,"* Report 9 (Revised), March 27, 2002.

Feldpausch, R. *"ESCAP II: E-Sample Erroneous Enumeration Analysis,"* Report 5, October 10, 2001.

Hogan, H., *"The Accuracy and Coverage Evaluation: Theory and Applications"* presented at the American Statistical Association Joint Statistical Meetings, 2000 Proceedings of the Section on Survey Research Methods, August 2000

Hogan, H., Whitford, D., *"Findings from the Census 2000 Accuracy and Coverage Evaluation,"* Invited Paper at 2001 American Statistical Association Joint Statistical Meetings, 2001 Proceedings of the Section on Survey Research Methods, August, 2001.

Krejsa, E. and Raglin, D., *"ESCAP II: Evaluation Results for Changes in A.C.E. Enumeration Status,"* Report 3, October 15, 2001.

Navarro, A. and Olson, D. *"Accuracy and Coverage Evaluation: Effect of Targeted Extended Search,"* DSSD CENSUS 2000 PROCEDURES AND OPERATIONS MEMORANDUM SERIES B-18*, February 28, 2001.

Mule, T., *"ESCAP II: Person Duplication in Census 2000,"* Report 20, October 11, 2001.

Mule, T., *"Revised Preliminary Estimates of Net Undercounts for Seven Race/Ethnicity Groupings,"* DSSD Revised A.C.E. Estimates Memorandum Series PP-2, April 4, 2002.

Robinson, J. G., "*ESCAP II: Demographic Analysis Results,"* Report 1, October 13, 2001.

Thompson, J., Waite, P., Fay, R., *"Basis of 'Revised Early Approximations' of Undercounts Released Oct. 17, 2001,"* document dated October 26, 2001.

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.