

## **Analysis Plan for Further ESCAP Deliberations Regarding the Adjustment of Census 2000 Data for Future Uses**

### **Background**

On March 1, 2001, The Census Bureau issued the Executive Steering Committee for A.C.E. Policy (ESCAP) recommendation that the Census 2000 Redistricting Data not be adjusted based on the Accuracy and Coverage Evaluation (A.C.E.) program data. The ESCAP was unable to conclude, based on information available at the time, that the adjusted Census 2000 data were more accurate for redistricting.

By mid-October, the Census Bureau will recommend whether Census 2000 data should be adjusted for future uses, such as the census long form data products, post-censal population estimates and Census Bureau demographic survey controls. In order to inform this decision, further research will be conducted generating data for ESCAP's review. The analyses will focus on resolving the concerns that ESCAP identified during its deliberations for the redistricting adjustment decision. This document describes the research agenda and is organized by the topic areas of concern.

The broad, overarching concern was that the Demographic Analysis and the A.C.E. estimates of the population were inconsistent. Even though alternative demographic estimates were produced by varying the assumptions underlying the Demographic Analysis, the highest reasonable estimate indicated that Census 2000 undercounted the population by 0.32 percent, while the A.C.E. produced a net undercount estimate of 1.15 percent<sup>1</sup>. In previous censuses since 1960, the Demographic Analysis estimates were used to evaluate decennial census coverage. The estimate derived through the 1990 coverage measurement survey was reasonably consistent with the 1990 Demographic Analysis estimate of the total population. When the corresponding estimates for Census 2000 were found to reflect substantial differences in the population estimates, this concerned the ESCAP. Four scenarios were identified that could explain this result:

- The 1990 census coverage measurement survey (Post Enumeration Survey), 1990 Demographic Analysis estimates, and the 1990 census may have understated the Nation's population, while Census 2000 included portions of this previously unidentified population.
- Demographic Analysis estimates might not have captured the full growth between 1990 and 2000, specifically due to static assumptions about critical components of international migration such as unauthorized migration, temporary migration, and emigration.

---

<sup>1</sup>The 1.15 percent and 0.32 percent of the undercount rates are based on census counts that include both the housing unit and group quarters populations.

- Census 2000, as adjusted by the A.C.E., might overestimate the Nation's population. This situation raises the possibility of an undiscovered problem with the A.C.E. or Census 2000 methodology.
- A combination of these explanations.

To address these possibilities, further research is required into the quality of the three independent measures of the population - the Demographic Analysis estimate, the A.C.E. estimate and the census count itself. Specifically, research will address whether the Demographic Analysis estimate was too low and/or whether the adjusted estimate was too high. The latter situation could have occurred if either the A.C.E. did not measure the coverage error accurately or the census count had coverage error reflected by components not measured by the A.C.E.

In addition, the ESCAP was concerned about two other issues related to the A.C.E. estimates - balancing error and synthetic error. Balancing error occurs in the A.C.E. when cases are handled differently in the two independent samples (the P- and E- samples) when identifying gross omissions and erroneous enumerations. This is explained more fully under section B.1.a below. Synthetic error reflects the extent that net census coverage within a post-stratum is not relatively uniform. Uniformity of coverage is the underlying assumption of the synthetic estimation process of carrying coverage correction factors down to the block level. The concerns regarding synthetic error are described more fully in section D below.

The analysis agenda is organized around four basic areas of research: 1) recalculation of Demographic Analysis estimates using new migration assumptions as well as new birth and death data, 2) A.C.E. issues, including balancing error, 3) Census 2000 issues and 4) synthetic error.

## **A. Demographic Analysis (DA) Research**

This area of research addresses the discrepancy of the demographic analysis data and the A.C.E. adjusted estimates of population. Specifically, this area of research will reexamine the historic levels of the components of population change to address the scenarios dealing with the possibility that the 1990 Demographic Analysis estimates understated the Nation's population and that demographic analysis did not capture the full growth between 1990 and 2000. Consultation with demographic experts inside and outside the Census Bureau has led to a research program consisting of a variety of research projects focused on the methodologies and underlying estimates of the components of population change. The research activities are concentrated in two areas:

## 1. International Migration

Assumptions regarding international migration are the most uncertain component of the demographic analysis estimates. The international migration component represents a combination of several components. Some of these components, e.g. legal immigration, are measured through continuous administrative data. For other components, e.g. temporary migration, emigration, and unauthorized migration, we do not have administrative data to provide continuous and current measurements. In the past, we have relied upon the most recent decennial data to develop a once a decade measure of these components. Thus, for the 1990 to 2000 decade, we would have relied upon the measurement from the 1990 census to develop an estimate for the 1990 to 2000 decade.

This work will involve examining preliminary data from the Census 2000 long form and the Census 2000 Supplementary Survey (C2SS) to provide information to update the measurement of the international migration components. Although the research will focus primarily on those components less well measured, e.g. emigration, temporary migration, and unauthorized immigration, the work will also include research into all of the current assumptions relating to the components of international migration. The first goal is to validate for the 1990 to 2000 period, the calculation of the components of international migration used in previous estimates. Then, using the preliminary data from the Census 2000 long form and possibly the Census 2000 Supplementary Survey (C2SS), we will develop some updated measures of the components of international migration. The second goal is to assess if the documented calculation of the 1990 to 2000 migration components affect the DA estimate for 2000 and thus account for some of the discrepancy with the A.C.E. results. Research to be conducted includes the following:

- We will examine the assumptions about international migration flows, specifically for unauthorized migration, legal immigration, emigration, temporary migration, and migration from Puerto Rico. Utilizing preliminary long form data from Census 2000 and other information sources (including C2SS), we can prepare the first set of documentation for our current international migration assumptions and we can assess the accuracy of assuming a continuation of the estimates developed from the 1990 Census data. Specifically, we will estimate migration using available long-form data on place of birth, citizenship, and year of entry and compare this estimate to the estimates previously used that were developed from the 1990 Census long form data. Thus we will evaluate differences in size and characteristics of previously implied flows based on current data sets. If appropriate, we will recalculate the demographic analysis estimates for 2000 employing any revised levels of international migration.

- We will assess the quality of the foreign-born and Hispanic population data (important because these data are major inputs to the setting of assumptions noted above). We will review edit and allocation procedures for foreign-born and Hispanic populations in the 1990 and 2000 censuses and attempt to quantify the effect (or at least address the direction of the effect) of any differences. We also will review the impact of any change in the edits and allocation procedures on the size and characteristics of these population groups.

## 2. Robustness of Demographic Analysis

In addition to the research aimed at examining the components of international migration used in the demographic analysis estimates, we will examine the remaining assumptions underlying the Demographic Analysis components of change. These components include the birth, death, and Medicare components. This work will entail the following:

- We will examine the consistency of the components by cohort and age/sex groups across time (1935 to 2000), including the historical international migration components. We will construct DA undercount rates for the 1940 to 2000 decennial censuses and examine them for consistency. We will examine the consistency of sex ratios across cohorts and age/sex groups. Inconsistent or anomalous results will be noted, and possible reasons identified.
- We will review the assumptions about the completeness of vital statistics registration. Specifically, we will review the historic levels of births and deaths used to develop existing DA estimates and the assumptions about the underregistration of births and registration of infant deaths. We will evaluate both the procedures for adjusting births for underregistration and the level of historical deaths (both total and by age). If appropriate, we will redevelop the historical annual levels of births and deaths to 1990 and 2000.
- We will examine the assumptions about the variation and coverage of Medicare data. This work will include documenting the differences in the sources of Medicare data used in the 1990 and 2000 DA estimates, evaluating the adjustment rates used for underenrollment in the 1990 and 2000 DA estimates, and reconciling the differences in the Medicare files for 1990 and 2000.
- If appropriate, we will recalculate the demographic analysis estimates for 1990, compare them to the original 1990 Demographic Analysis estimates, and assess their impact on the DA estimates for 2000.

- We will analyze the consistency of DA estimates of the population, by race, ethnicity, and nativity, with Census 2000 and A.C.E. This work will entail 1) developing DA benchmarks of the population, by selected race, ethnicity, and nativity groups, 2) obtaining census tabulations of the native and foreign-born populations from preliminary Census 2000 and the 1990 Census long forms, and 3) comparing to the DA benchmarks to derive coverage estimates by selected age, sex, and race groups.

## B. A.C.E. Issues and Planned Research

### 1. Major Areas of Research

#### a. Balancing Error

The A.C.E. was conducted using a defined area of search, the sample blocks and surrounding blocks for clusters selected for targeted extended search. There were concerns, since there was a change in the 1990 procedure of expanding the search area to surrounding blocks for all sample blocks. We found 3 million more matches in surrounding blocks than correct enumerations after expanding the search area. This difference must be explained in terms of its impact on subsequent estimates of total population. There are two scenarios:

- C The unit is located in the surrounding block with no effect on estimates of coverage, but would explain the three million difference.
- C The unit is outside the search area and the corresponding people should have been coded erroneous enumerations. This would result in an overestimate of the net undercount.

This may have been compounded by the **targeting** used in the A.C.E. to match in an area of search around the sample blocks, i.e., the search area. This targeting to make searching effective may have introduced limitations and/or biases into our measurement of coverage. There were three specific concerns in our review of the 2000 A.C.E.

- C There were a number of census people that might have been coded as correctly enumerated although the housing unit was not actually located in the sample block. If we didn't estimate the correct number of erroneously enumerated cases, the result would be an overestimate of the net undercount.

C The P-sample may have incorrectly included some housing units in a neighboring block, then in the extended search, the people would have been recorded as matching to the census in the surrounding blocks. Hence, these cases would appear to be balancing error when, in fact, the extended search was compensating for the original listing error. If the P sample had more geocoding error than expected, the Targeted Extended Search (TES) would have compensated for the error and the impact would be trivial and would have little or no impact on final coverage estimates. This would help explain some of the differences of the apparent lack of balance of 3 million.<sup>2</sup>

C Problems in identifying census geocoding errors may have affected the sampling used to select people for extended search outside the sample blocks. That is, the TES sample could have excluded cases it should have included and thus, not matched or followed up on them correctly. The effect of their exclusion would be an overestimate of the net undercount.

It is likely that all of these errors occurred to some extent. What is not yet known is the scale of the error and whether the magnitude of the error was such as to significantly affect the relative accuracy of the A.C.E. adjusted numbers. The additional geographic field work is described in more detail in the attachment.

b. Erroneous Enumerations

Subsequent to the March 1<sup>st</sup> decision, a new area of concern was identified. In comparing the A.C.E. measures to the comparable measures from the 1990 Census, the Census 2000 erroneous enumerations were found to differ substantially from the 1990 measures. These differences indicate concerns that the level of erroneous enumerations may be understated for Census 2000. Therefore, these differences must be explained because an understatement of erroneous enumerations results in an overstatement of net undercount. Research described below will quantify the accuracy of the A.C.E. measures of erroneous enumeration.

---

<sup>2</sup> Assume 2.6 million of the P-sample are listed in the surrounding blocks. If 95% of them are in the search area (a plausible percentage), and if 90% match (about the overall match rate), then we have accounted for 2.2 million matches to the surrounding blocks. When we divide this 2.2 million by the P-sample coverage of 0.94, we have accounted for about 2.36 million of the 3 million lack of balance.

- C The Analysis of Measurement Error Study will determine how well the A.C.E. identified erroneous enumerations and correct enumerations. This study is based on a reinterview of a sample of E-sample records. This is described more fully in section B.1.c below.
- C Another evaluation based on results from the ‘E-sample Erroneous Enumeration Study’ will analyze the erroneous enumerations for various characteristics. This evaluation will compare the rates of the different types of erroneous enumerations for Census 2000 with corresponding 1990 rates. This evaluation will also recategorize people with unresolved status into the appropriate erroneous enumeration categories by using data from the followup forms. The goal of this work will identify explanations for differences between 1990 and 2000 coding of erroneous enumerations.
- C The duplication study discussed in Section C1 will also provide information regarding the differences between 1990 and 2000. This study will validate whether the A.C.E. process is correctly coding census 2000 duplicate enumerations as erroneous.

c. Total Error Model and Loss Functions

Loss function analyses, reviewed by the ESCAP during its deliberations on whether to adjust the census redistricting data, were based on a Total Error Model that corrected the A.C.E. for biases, thus producing measures of the “true” population that could be used as standards for comparing the adjusted and unadjusted census results. The 1990 Total Error Model was adapted to the extent possible to “fit” the 1990 coverage measurement survey error components into the 2000 survey design. This model was updated with available Census 2000 data, but retained several error component measures obtained from the 1990 coverage measurement survey and 1990 evaluations, because the 2000 A.C.E. evaluation data were not yet available. Thus, the error model assumed that the actual A.C.E. error rates for these components were similar to those reflected by the 1990 coverage measurement survey results. This was viewed as conservative because it was expected that the A.C.E. was of higher quality than the 1990 coverage measurement survey. Work is underway to validate that the assumption above is correct.

We are conducting studies to revise the 1990 Total Error Model to reflect actual A.C.E. error components, as measured by 2000 evaluations. Because of methodological changes between 1990 and 2000, there are issues that

influence the comparability of this updated analysis to the March 2001 analysis. The analysis will include a discussion of the comparability.

The A.C.E. error components that were previously based on 1990 data will now be measured and input into the revised Total Error Model are:

- P-sample matching error
- P-sample data collection error
- P-sample discrepancy error
- E-sample processing and data collection errors

Synthetic error is not included in the total error model - this component of error is discussed later. A.C.E. error rates for these Total Error Model components will be obtained from the following evaluation studies.

- The *Matching Error Study* will provide the A.C.E. P-sample matching error rate and E-sample processing error rates. The methodology consists of the clerical rematching of all of the people in a one-fifth subsample of the A.C.E. clusters by expert matchers to determine the best match code possible. We will compare that match and residence information to the production codes.
- The *Analysis of Measurement Error Study* uses the results of the *Evaluation Followup Interview* to provide the error components for E-sample and P-sample data collection error relating to person coverage, and P-sample discrepancy error. The methodology consists of revisiting some of the households in a one-fifth subsample of the A.C.E. clusters and using that information to rematch the Census and A.C.E. people in those households. The results of this study will determine the accuracy of the data going into the person matching process, such as the results from Census and A.C.E. questionnaires. This can involve reclassification of correct and erroneous enumerations. We will determine the accuracy of the residence status of A.C.E. people and how well the A.C.E. process identified Census erroneous enumerations (EEs) and correct enumerations (CEs).

Once the Total Error Model is updated with current data, new loss function analyses will be conducted. The loss function analyses will be expanded to analyze the accuracy of governmental units, in addition to states and counties.



No loss function analyses will be run for congressional districts.

d. Correlation Bias

*Correlation bias* in Dual System Estimates (DSEs) results from a failure of the general independence assumption underlying DSEs due either to causal dependence or heterogeneity. Causal dependence occurs when the act of being included in the census makes someone more likely or less likely to be included in the A.C.E. Heterogeneity occurs when the census and A.C.E. inclusion probabilities vary over persons within post-strata. When heterogeneity within post-strata exists it is generally suspected to be of the form where persons more likely to be missed in the Census are also more likely to be missed in the coverage survey (A.C.E.). This will lead to underestimation of true population by the DSEs. The direction of the effect of causal dependence, if it exists, is less certain.

Correlation bias in the A.C.E. estimates, whether due to heterogeneity or causal dependence, was assessed by comparing A.C.E. and DA results. Correlation bias estimates available for the March 1, 2001 ESCAP recommendation used DA estimates as of February 26, 2001. If further DA research results in revisions to the DA estimates, then the correlation bias estimates will be recomputed. The revised correlation bias estimates will then be used as inputs for revisions of the total error model and loss function analyses.

2. Auxiliary Areas of Research

This section describes other areas that did not preclude ESCAP from recommending that Census 2000 data should be adjusted for redistricting purposes, but for which ESCAP would have preferred additional data. Further research in these areas will be conducted in order to confirm the ESCAP's conclusions.

a. Missing Data

Missing data occurs in the A.C.E. if after all followup attempts there remain households that were not interviewed or households with some portions of the person data missing such as age or race. Sometimes the missing item involves the status of whether a person matched, was a resident on Census day or was correctly enumerated.

For a small number of people in the P-Sample, there was not enough

information available to determine the match status (whether or not the person matched to someone in the census in the appropriate search area) or the resident status (whether or not the person was living in the block cluster on Census Day). Determining residence status was important for the P-Sample because Census Day residents of the block clusters in the sample were used to estimate the proportion of the population who were not counted in the census. Similarly, some people in the E-Sample lacked information to determine whether the person was correctly enumerated. Generally for cases with missing status a probability of resident, match, or correct enumeration was assigned based on information available about the specific case and about cases with similar characteristics.

The rates of occurrence of unresolved A.C.E. cases for match status, correct enumeration status, and mover status were viewed as low enough to preclude serious biases in the A.C.E. results. We are now doing analysis of the missing data model to determine if the assumptions are correct.

We will develop and apply alternative models for the treatment of missing data. These alternative models will be carried through A.C.E. estimation process so that the effect on DSEs can be assessed.

b. Late Census 2000 Additions

The levels of late Census 2000 additions were significantly higher than in the 1990 census. Late additions are those persons included in the final census counts, but which due to their late inclusion were excluded from in the A.C.E. matching and dual system estimation processes. For Census 2000, the late additions consisted exclusively of housing units that were temporarily removed from the census because they were suspected to duplicate other housing units, but which were later (after the A.C.E. matching process started) reinstated into the final census after further research was conducted. This differs from the 1990 Census in which the late additions were persons who were enumerated too late in the census cycle to be included in the matching and dual system estimation processes and were not factored into the coverage ratios. The A.C.E. design treated the late census data appropriately in measuring the census undercount. Two areas of concern require further investigation - whether calculating DSEs without these additions resulted in a bias in the estimates and whether these impacted the assumptions underlying the synthetic estimation model.

There is no expectation of a bias in the dual system estimate caused by excluding late additions. The dual system estimate can be expressed as a

product of the 1) number of A.C.E. people and 2) the ratio of census complete and correct enumerations to the number of people in both systems.

Consequently, any effect must come from one of these two sources. Excluding the late additions does not impact the estimate of the number of A.C.E. people, which come solely from the A.C.E. enumerated sample. Excluding the late additions also will not affect the dual system estimate of the true population if the number of matches is reduced proportionately to the number of census correct enumerations. Given the traditional dual system independence assumption, one would expect this result. Consequently, there is no expectation of a bias in the dual system estimate caused by excluding late additions. Data were not available at the time to validate this assumption.

We will now attempt to validate this assumption by performing a rematch of the P- and E-samples, with the late additions included in the E-sample, to attempt to measure the impact on the rates for correct enumerations and duplicates. This rematch will be conducted in a one-fifth subsample of A.C.E. clusters. This study has limitations because only computer and clerical matching can now be performed; that is, no field work will be conducted. Consequently, a high rate of unresolved cases is expected.

The concerns regarding synthetic error are addressed in Section D. "Synthetic Error".

c. Conditioning

Conditioning error occurs under two scenarios:

1. Census data collection affects the A.C.E. This will be measured in the correlation bias.
2. A.C.E data collection affects the census. This will be examined in the evaluation described below.

The effect of potential conditioning of Census 2000 respondents by the A.C.E. operations was assumed to be minimal, similar to the 1990 findings. The research is necessary to confirm this assumption.

An evaluation will examine whether census and A.C.E. operations were kept operationally independent. The analysis will be based on comparing Census 2000 results in A.C.E. and non-A.C.E. blocks.

d. Mover Status Analysis

The match rate portion of the DSE formula (M/P) uses persons with all types of mover status (nonmovers, outmovers, and in-movers), differentiating between the different types of mover status. Therefore, misclassification of mover status could cause the DSEs to be overstated, understated, or both, depending on the post-strata.

The *Measurement Error Reinterview Analysis* will measure the extent of mover misclassification by using the results from the *Evaluation Followup Interview*.

e. Housing Unit Coverage

The coverage of housing units will be available in the late summer of 2001. These data will be examined in relation to person coverage estimates for 2000. These data from 2000 will be compared to the 1990 estimates of person and housing unit coverage.

In addition, another study will assess the impact of housing unit coverage on person coverage. This study looks at the P-sample to analyze the effect of housing unit nonmatches on the person nonmatches. The E-sample is also examined to help understand the relationship of housing unit status to person status. The correctly enumerated people in erroneously enumerated housing units are of particular interest.

f. P-sample Nonmatch Analysis

The P-sample nonmatches are examined for variables such as race domain and age/sex group to see if the nonmatches are different for various types of people. This aids in the understanding of the components of A.C.E. and also helps explain the differences between A.C.E. and DA. In addition, the nonmatches from 2000 are compared to the nonmatches from 1990. In conjunction with the analysis of the E-sample, it helps explain the differences between 1990 and 2000.

**C. Census 2000 Issues and Planned Research**

Research will be conducted into two components of the census - duplication issues and imputation of persons. A high level of duplication not measured by the A.C.E. design could cause the adjusted census estimate to be too high. The effect of imputed persons records are

also not measured by the A.C.E. The number of person records that were imputed in Census 2000 was significantly higher than in the 1990 census. The assumption is that the imputed persons are no different than the persons included in the A.C.E. process and therefore match rates are not impacted.

1. Duplication Not Measured in A.C.E.

The A.C.E. methodology by design did not measure duplication between components of the population living in group quarters and in housing units because group quarters were outside the A.C.E. universe. The A.C.E. also did not measure duplication within the group quarters population. Significant duplication of these types could explain some of the differences between demographic analysis and the adjusted Census 2000 data.

The A.C.E. E-sample will be computer matched to the entire census to determine the extent of duplicate enumerations that were not in scope for the A.C.E. This analysis will potentially explain some of the differences between demographic analysis and the A.C.E.

We also plan an extended computer search within the A.C.E. E-sample for duplicate census enumerations among housing units and also between housing units and group quarters persons (which were out-of-scope for A.C.E.) This will help to explain differences between the A.C.E. and the 1990 coverage measurement survey.

2. Census Person Imputations

Census 2000 imputed a higher number of cases than in the 1990 census that came through the process with little or no information as to the occupancy status, or with an occupied status, but with no definitive population count. In addition, Census 2000 imputed more whole person records in cases with known household sizes, but with all the person data missing for some or all of the household members. Although the A.C.E. handled imputed persons appropriately in the estimation process, there was concern about not having information as to what census design processes contributed to the number of imputed persons when compared to the 1990 census.

Given the potential impact that this level of imputations may have on Census 2000 data, it is essential to understand the demographic characteristics of the imputed people and how this may help explain the difference between the census and demographic analysis, as well as, how the imputations affect differences between the E-sample in 1990 and the E-sample in 2000.

There were concerns expressed regarding the effect of whole household imputations on

the heterogeneity assumption but these concerns are studied under the synthetic error analysis in Section D.

## **D. Synthetic Error**

The synthetic assumption states that census net coverage does not vary within post-strata. For example, the synthetic assumption implies that census counts in Florida in a particular Hispanic post-stratum have the same net coverage as the census counts in the same Hispanic post-stratum but in New York. The synthetic assumption within post-strata will permit the Census Bureau to draw conclusions from the A.C.E. sample about the population as a whole and then apply them to individuals living in geographic areas smaller than post-strata. The synthetic assumption is necessary to permit correction for small geographic areas based on a sample. This adjustment is only correcting for systematic biases and not local census errors. The error that is introduced when the synthetic assumption does not hold is called synthetic error.

Synthetic estimation methodology is directed at correcting for a systematic under- or overcount in the census. The synthetic estimates will not result in the correction of random counting errors that occur for any entity (blocks tracts, counties, etc). Therefore, the synthetic estimate will not result in extreme changes in small geographic entities, nor will it correct for extreme errors. It is designed to remove the effects of systematic errors so that when small entities are aggregated, systematic and differential coverage errors are corrected.

In the assessment of accuracy, the Census Bureau is concerned with synthetic error since it is not included directly in the total error model. The analysis of the effects of synthetic error were based on the construction of “artificial populations.” These are populations that are created with surrogate variables that are known for the entire population, and are developed to reflect the distribution of net coverage error. This analysis of synthetic error and its effect on the loss functions was limited.

Our additional analysis will expand the scope of the earlier artificial population work and add an approach using direct estimates of coverage at lower geographic levels.

### **1. Using Artificial Populations**

We will do a sensitivity analysis on the results from B-14. B-14 gave results for weighted and unweighted loss functions using one of two methods for distributing targets to post-strata and one of 8 models for correlation bias and percent of 1990 processing bias. This work will concentrate on the weighted loss functions and analyze the sensitivity of the B-14 results over both the methods for distributing targets to post-strata and all 8 models. Once again this analysis will be conducted for states and congressional districts.

## 2. Using Direct Estimates

We will calculate direct DSEs for census divisions and for states having sufficient sample size to produce direct estimates with reasonably low variance. Assuming the resulting direct DSE population estimates are unbiased, the mean square error of the production synthetic estimate of total population will be estimated.

### **E. Schedule**

Some of the A.C.E. evaluation work being undertaken involves field work and/or additional computer or clerical matching work. The Evaluation Followup Interview was conducted in the field during the winter of 2001. The Matching Error Study matching work was completed in the spring. Results from these studies are being processed, with initial data being available for review in early summer. Field and clerical work for the TES2 and TES3 (described in the attachment) studies began in the winter and will continue into July. Results from these studies won't be available for ESCAP review until later in the summer. Matching for the late census adds evaluation is scheduled for late-July, with data available for review in August. Other research is being conducted on a flow basis as data become available and analyses are conducted.

The ESCAP began holding weekly (or more frequent) meetings to review analyses of data related to the topics of concern beginning on June 18. It is expected that all of the research and analyses described will be completed by the end of September. The ESCAP will then discuss how the results impact their concerns and will make a recommendation by mid-October as to whether adjusted or non-adjusted census data should be used for subsequent purposes.

During the September through October time frame, analysts will document the results of their research in evaluation reports, finalizing them in time for release to the public concurrently with the ESCAP recommendation.

### Field Operations to answer the concerns about lack of balance

In order to answer these concerns and explain the lack of balance present due to Targeted Extended Search (TES) and to explain the lack of balance that may be introduced due to TES, we will be examining the results of Targeted Extended Search 2 (TES2) and Targeted Extended Search 3 (TES3). TES2 followed up E-sample housing units that were coded as erroneous enumerations in the initial housing unit phase to determine if the unit was inside or outside the block cluster and surrounding rings. TES3 will follow up other types of units, both P-sample and E-sample, that may contribute to a lack of balance.

In TES2 we are evaluating the housing units coded during the housing unit matching as not existing as housing units within the cluster. The block containing the housing unit selected for additional geographic work and the surrounding blocks were identified on a map. The field representative identified the block where the housing unit existed and the housing unit was classified as:

- C Existing in the surrounding blocks
- C Existing outside the surrounding blocks
- C Existing within the block cluster
- C Not a housing unit
- C Unresolved

So, a housing unit may be coded as in surrounding blocks or outside the search area when it was part of the block cluster.

In TES3 we are also sending to the field a sample of census housing units classified as correctly enumerated in the block cluster. If a housing unit was classified as correctly enumerated in the block cluster in error, the housing unit was not eligible for targeted extended search in person matching. This could explain more of the lack of balance identified in the person matching.

In addition, we are sending additional types of P-sample cases for more geographic field work and a sample of matches in the sample block as a control. These types of cases are:

- C P-sample people matched in surrounding blocks
- C Not matched P-sample housing units
- C P-sample people matched in the sample block cluster