ASSESSMENT OF ACCURACY OF ADJUSTED VERSUS UNADJUSTED 1990 CENSUS BASE FOR USE IN INTERCENSAL ESTIMATES

REPORT OF THE COMMITTEE ON ADJUSTMENT OF POSTCENSAL ESTIMATES
BUREAU OF THE CENSUS
DEPARTMENT OF COMMERCE
AUGUST 7, 1992

RECOMMENDATION

The Committee on Adjustment of Postcensal Estimates (with an acronym of C.A.P.E. but referred to in this report as the Committee) investigating potential census adjustment for intercensal population estimates concluded that on average, an adjustment to the 1990 base at the national and state levels for use in intercensal estimates would lead to an improvement in the accuracy of the intercensal estimates. (Attachment 1 contains a list of the members of the Committee.) This conclusion was based on a set of extensive research and analyses as well as input from outside consultants. This outside technical advice included a Panel of Experts whose work culminated in a day-long meeting with Census Bureau staff. (Attachment 2 contains a list of the Panel of Experts.) Under the auspices of the Office of Management and Budget (OMB), there also was consultation with other Federal agencies, which are prime users of intercensal estimates.

In coming to its conclusion, the Committee did not vote. Instead, there was an attempt to reach consensus. The conclusion of the Committee was not unanimous, but the large majority of the Committee agreed with the finding. Since there was no vote, this report does not contain a specific listing of minority opinions. Rather, a series of concerns is listed. There was general consensus on several key points.

1. This decision was separate and distinct from the June 1991 decision about whether to adjust the 1990 census for all uses. Making a decision about whether to adjust the full census is quite different from deciding whether to adjust the base that is used in mathematical algorithms to produce estimates of population at several points in the decade between censuses (intercensal estimates).

2. The majority of the Committee concluded that on average, an adjusted state base would be more accurate than an unadjusted state base for use in intercensal estimates, but the Committee recognized there is not necessarily improvement for each and every state base. In fact, the Committee was concerned about a few specific states where the evidence was inconsistent as to whether adjustment was making an improvement. Even so, the Committee felt that overall there was improvement at the state level.

3. States are an important political entity and the first tier in most funding programs. Therefore, the Committee felt that every state or none of the states should be adjusted. Even though some states are smaller than several large cities, the Committee did not recommend adjusting selected cities or counties.

4. For smaller areas (generally, areas of less than 100,000 population), some of the Committee judged that the use of an unadjusted base for the estimates was better than the use of an adjusted base. Other Committee members concluded there was no way to determine whether an adjusted or unadjusted base was more accurate. In the absence of data showing improvement by adjustment, the Committee concluded that the relative distribution of population by substate areas within each state was more

accurate using census counts than the comparable relative distribution using adjusted counts.

5. The Committee was quite concerned about adjusting some, but not all substate areas, especially since there was no way to determine the cutoff of which areas to adjust and there had been no research on the effect of adjustment for a partial set of substate areas.

The Committee's technical assessment was based on a massive amount of data. While there was a re-examination of the information already collected in conjunction with the evaluation of the Post Enumeration Survey (PES), the Committee relied mostly on a large volume of additional research conducted since July 1991. In performing this additional research, the Census Bureau had more time so it could take full advantage of what it had learned from its analysis to date of the 1990 census and the PES. The Census Bureau also had fewer constraints to use prespecified procedures compared to the process in conjunction with the July 1991 decision whether to adjust the 1990 census for which a court order required prespecified procedures. This additional research turned out to be extremely useful, not only for this decision, but for future surveys of all kinds, including those designed for potential adjustment. The Committee wants to acknowledge specifically the massive effort that the professional statistical staff at the Census Bureau put into this research. It was research of such quality that all those involved should be rightly proud. The quality and usefulness of the research also were noted by the set of outside experts that helped review Census Bureau research.

A full description of this research is beyond the scope of this report, but a summary is provided. There are, however, extensive minutes of the Committee meetings, which contain, as attachments, the major results of the additional research. The Committee would like to commend David Whitford and Michael Batutis for preparing these excellent minutes.

In addition to providing useful information, this additional research detected some errors and made some refinements to the levels of estimated undercount originally reported in the spring of 1991. These changes are summarized in the following table and described more fully later in the report.

Population Group	Estimated Undercount			
	June 1991		July 1992	
	Undercount Estimate	Sampling Error	Undercount Estimate	Sampling Error
U.S. Total	2.08%	.18%	1.58%	.19%
Black	4.82	.29	4.43	.51
Asian and Pacific Islander	3.08	.47	2.33	1.35
American Indian, Eskimo, or Aleut	4.77	1.04	4.52	1.22
Hispanic (Can be of any race)	5.24	.42	4.96	.73

This report is a summary of the process that led to the Committee's recommendation. Though the report concentrates on activities that took place late in the decision process, the report also covers several topics that were discussed throughout the year of deliberations by the Committee. Some readers of this report may desire further background on the issue of undercount in a census and the efforts of the Census Bureau to measure and potentially correct (adjust) for any such undercount. There are numerous documents that could be read for background. One good summary document is the notice in the Federal Register concerning the decision of the Secretary of Commerce about whether to adjust the 1990 census (Reference: Federal Register, Volume 56, #140, Part III, pages 33582-33692). The remainder of this report is divided into several sections.

BACKGROUND - UNDERCOUNT — This section contains a description of coverage in the decennial census as well as the methods the Census Bureau uses to measure coverage.

BACKGROUND - ESTIMATES — This section contains a description of why the Census Bureau undertook the task of examining whether to adjust intercensal estimates as well as a very brief description of the estimates program and its use.

RESEARCH — This section summarizes the additional research done since July 1991. This research was the major foundation for the Committee's assessment.

DECISION — This section briefly describes the decision process of the Committee as well as the Executive Staff. These final discussion as well as the year long deliberations of the Committee will be key pieces of input to the Director's decision.

FUTURE — This section contains a few general findings concerning the process of measuring undercount in the future.

BACKGROUND ON UNDERCOUNT

The issue facing the Committee was whether potential error in the PES and adjustment technology was at a sufficiently low level to recommend the inclusion of results from the PES into intercensal estimates. The decennial census is also subject to error, and the PES tries to measure the net coverage error in the census.

This section describes the operations of the 1990 PES to measure census coverage error and how these PES results might have been used for a potential adjustment of the 1990 census. This section is provided solely for background, so the section can be skipped for those already familiar with coverage error in a census as well as the Census Bureau's methods to measure coverage error by the PES and Demographic Analysis.

Since the very first census, there have been problems in accurately counting every person living in the United States. The resulting undercount, or percentage of the population that is not counted by the census, is not a new phenomenon. Beginning with the 1940 census, each decennial census has included an evaluation program to attempt to measure the extent of undercount, or what is often called coverage error. These evaluations showed a steady improvement in net census coverage over four decades, from an estimated undercount of more than 5 percent for the total population in 1940 to an estimated undercount in 1980 of just over 1 percent. They also have shown larger undercount rates for the Black population than the non-Black population and a differential that has stayed about 3-4 percentage points over the period. A difference in estimated undercount for one population subgroup (like Blacks) and another population subgroup (like non-Blacks) is called the differential undercount.

Because of concern about this differential undercount, it was suggested that if the Census Bureau can estimate the number of people missed in a census, why not simply correct the census to account for missed persons and thereby make the census more accurate. This, in simple terms, is what is called "adjustment." But estimating the census undercount with acceptably small error and, in turn, using that knowledge to improve the census counts for all levels of geography are two highly complex and difficult tasks.

The Census Bureau had two major programs to measure coverage in the 1990 census. The first was the PES, which was a sample survey taken after the census. Approximately 165,000 housing units in a sample of 5,290 census blocks or block clusters were interviewed. Block clusters are combinations of small blocks. For the rest of this report, block will be used to mean a block or a block cluster. Persons enumerated during the PES were also referred to as the P-sample. After persons in the housing units in the selected sample blocks were interviewed, their responses were matched to census records in the same set of blocks to determine whether they were counted in the census. This process measured erroneous omissions in the census.

The Census Bureau also measured erroneous inclusions in the census by determining whether any of the persons in the PES sample blocks who were enumerated in the census should not have been counted or should not have been counted at that particular location. An erroneous census enumeration, for example, could have included a child born after April 1, 1990, a person who died before April 1, or a college student away from home who was enumerated at his or her parents' address, instead of being correctly enumerated at the college. Persons in this sample constitute the E-sample.

The data on erroneous inclusions and erroneous omissions were used to produce an estimate of the net undercount or net overcount of the population in the census. This was a very complex process that combined elements of survey design, interviewing, matching, imputation, mathematical modeling and professional judgment.

Second, the Census Bureau used a system called Demographic Analysis (DA) to also measure census coverage. Basically, in DA, an independent estimate of the total population is produced by combining various sources of administrative data. This process included using historical data on births, deaths, and legal immigration; estimates of emigration and undocumented immigration; and Medicare data.

Demographic analysis estimates were used to evaluate the reasonableness of the PES estimates. Only the PES provided estimates of undercount and overcount at a level of detail suitable for use in potential adjustment. For example, demographic analysis estimates were produced only at the national level and for the Black and non-Black populations; the PES process was designed to measure coverage error for more population subgroups (Whites, Blacks, Hispanics, Asians and Pacific Islanders, and American Indians) by detailed levels of geography. Therefore, only the PES data could permit an adjustment.

Each of these programs will be summarized below. For a more detailed discussion of PES see Howard Hogan, "The 1990 Post-Enumeration Survey: An Overview," a paper presented at the American Statistical Association in August 1990; for a more detailed discussion of Demographic Analysis see J. Gregory Robinson, "Plans for Estimating Coverage of the 1990 United States Census: Demographic Analysis," a paper presented to the Southern Demographic Association, in October, 1989.

POST-ENUMERATION SURVEY (PES)

Sample Design

The PES sample was selected in stages. First a random sample of blocks was drawn. Blocks are small polygons of land surrounded by visible features. Most are like the four-sided blocks in a city. Within the selected set of sample blocks, all housing units were listed.

To select the sample of blocks, all blocks in the United States were assigned to one of 101 groups called strata. The strata were defined by geography, city size, racial composition, and percent of housing units that were renter occupied as opposed to owned. A representative sample of blocks was selected from each of the sampling strata. A separate sampling stratum was defined for American Indian Reservations.

Persons living in institutions were excluded from the PES, as were military personnel living in barracks, people living in remote rural Alaska, and persons in emergency shelters and persons who had no formal shelter.

Listing and Interviewing

In February 1990, Census Bureau interviewers who are part of the permanent Census Bureau staff of interviewers visited each of the sample blocks to list all housing units. To preserve independence, none of the temporary enumerators hired to take the 1990 census was used for this listing operation and the listing operation was not conducted out of the temporary census offices. The reason for this was to make sure that temporary people taking the census did not know where a PES sample block was, because if they did, that block might be treated differently during the census.

After the completion of the regular 1990 census interviews, PES interviewers interviewed persons at households in the PES sample blocks. Although this interviewing drew from interviewers who had already worked on the 1990 census, steps were taken to preserve independence, such as not allowing an interviewer to work in a block in the PES that he or she had worked in during the census.

During the PES interview, the interviewers determined who was living in each housing unit, obtained their characteristics, and asked where they lived on April 1, 1990, Census Day. This latter question was necessary in order to determine whether those people who had moved since census day had been counted in the census. The PES interviewing began nearly 3 months after Census Day.

There was a quality assurance program for the interviewing phase to ensure that the interviewers really visited the household and that the people listed were indeed real. If interviewers made up people, they would not match to the census and would inflate the undercount rate.

Matching

The next step was to match the persons enumerated during the PES (the P-sample) to the census. Those persons in the P-sample matched to the census were considered to have been counted in the census; those nonmatched were considered to have been missed.

Matching was carried out in several stages. It involved an initial stage of computer matching followed by clerical matching to attempt to resolve cases that the computer could not match. Many of the persons not matched to the census by computer and clerical matching were assigned for a follow-up interview, if it was determined that additional information might help establish whether a match to the census was appropriate. An additional stage of clerical matching was then conducted using the information from the follow-up interview.

The E-sample, those persons in the PES blocks who were enumerated in the census, was examined to determine if they were correctly enumerated. E-sample persons were matched back into the census to determine if they were enumerated more than once (duplicates). The E-sample persons who were not matched to the

P-sample were potential candidates for erroneous enumerations. Some of these unmatched census persons were also included in the PES follow-up operation described above.

A final matching and reconciliation operation took place at the conclusion of the PES follow-up. An important aspect of this operation was that situations arose where correct match status for persons in the P-sample, or correct enumeration status for persons in the E-sample, could not be determined. This situation occurred because the initial interview was inconclusive or because an incomplete interview was obtained during the follow-up.

Imputation and Dual System Estimates

A final PES computer file was created that reflected the match status for persons in the P-sample and the enumeration status (correct or erroneous) for persons in the E-sample. Computer editing or imputation was performed to correct, insofar as possible, for missing or contradictory data. A critical aspect of imputation involved the estimation of a final match status for those persons whose match status could not otherwise be resolved.

The data in the final PES file were then summarized and incorporated with data from the full census to produce dual system estimates (DSE's) of total population. Dual system refers to the fact that two systems (the census and the PES) are used to make the population estimate. The DSE's were produced separately for each of 1,392 unique subgroupings of the population called post-strata. (See the following section titled Post-strata)

The DSE model to estimate total population conceptualized each person as either in or out of the census cross classified as either in or out of the PES. Essentially it involves determining how many people were (1) in the PES and in the census(matches), (2) in the PES and out of the census(Non-matches), (3) in the census but not in the PES, and (4) in neither the census or PES.

To get an estimate of total population, you could add up the four cells listed above. But, only two of those were directly estimated (cell 1, matches, and cell 2, non-matches). Making some assumptions and using some basic algebra, total population can be estimated without direct estimates for each of the four cells. These operations and the DSE are explained more fully in the Hogan paper cited above.

Post-Strata

The Census Bureau prepared the dual system estimates of the total population for each of 1,392 groupings of people called post-strata. The reason for forming the post-strata was to group persons who had similar chances (probability) of being counted in the census. A person's likelihood of being counted in the census (or in the PES) is called capture probability. The post-strata were defined by census division, geographic subdivisions such as central cities of large metropolitan statistical areas, whether the person was the owner or renter of the housing unit, race, age, and sex. Each person in the PES sample belonged in one of the unique post-strata.

For purposes of illustration, the following are examples of the 1,392 post-strata. One example is a post-stratum which contains Black males, age 20-29, living in rented housing in central cities in the New York primary metropolitan statistical area. A second example is that which contains non-Black non-Hispanic females, age 45-64, living in owned or rented housing in a non-metropolitan place of 10,000 or more population in the Mountain Division. A third example is that which contains Asian males, age 45-64, living in owned or rented housing in metropolitan statistical areas but not in a central city in the Pacific Division. A fourth example is that which contains non-Black Hispanic females, age 30-44, living in owned or rented housing in central cities in the Los Angeles-Long Beach primary metropolitan statistical area or other central cities in metropolitan statistical areas in the Pacific region. As can be seen from these examples, the 1,392 post-strata are very specific.

Adjustment Factors

The next step in the process was to compare the estimated total population for each post-stratum (the dual system estimate or DSE) to the census count to determine a "raw" adjustment factor. For example, if the DSE for a particular post-stratum was 1,050,000 and the census count was 1,000,000, then the adjustment factor was 1.05, reflecting about a 5 percent estimated net undercount. Though most adjustment factors are larger than one, indicating an estimated undercount, an adjustment factor may be less than one, which would have the effect of lowering the census count for the post-stratum if an adjustment is applied. This situation results when there is evidence of an overcount in the post-stratum.

"Smoothing" the Adjustment Factors

The next step was "smoothing" these "raw" adjustment factors to reduce sampling variance and to produce final adjustment factors. Because the PES was a sample, it was subject to sampling error. Sampling error is the error associated with taking some of the population (a sample) rather than all of the population (a census). The process of smoothing the "raw" adjustment factors to create final adjustment factors was a step to minimize the effect of sampling error. Basically, smoothing is a regression prediction model. A multi-variate regression using items correlated with undercount predicts the undercount for each of the 1,392 post-strata. Then, the final adjustment factor is an average of the "raw" adjustment factor and the predicted adjustment factor. For a post-stratum with low estimated sampling error, there was heavy weight on the "raw" adjustment factor in the averaging, and vice versa. The smoothing technique was based on certain assumptions and would add an additional component of error called model error. The Census Bureau hoped that the reduction in sampling error from smoothing would offset any additional errors from the smoothing model chosen. If the Census Bureau had not used smoothing, the final adjustment factors for some of the post-strata would have been based on estimates of undercount that were subject to very large sampling error.

Small Area Estimation

The Census Bureau used the final adjustment factors to produce adjusted counts for every block in the Nation. The PES can only produce "direct" estimates of the total population for relatively large geographic areas (*i.e.*, the 1,392 post-strata). If there had been a decision to adjust, however, the adjustment would have been applied to each of the Nation's approximately 5 million populated blocks. The Census Bureau developed a model that took the adjustment factors produced for each of the 1,392 post-strata areas and used them to estimate adjustment counts for each block. Since each of the post-strata contain many blocks parts, the Census Bureau based its model on a critical assumption that coverage error is similar for all blocks parts within a post-stratum. (A block part is simply that part of the block that falls within the definition of a post-stratum. For example, females within a block would be part of a block and in one set of post-strata while males within a block would be in different set of post-strata.) This assumption of all block parts within a post-stratum being alike (homogenous) with regard to the chance of being counted is analogous to the homogeneity assumption for persons.

Finally, the Census Bureau produced a set of census tabulations with adjusted counts. It did this by adding or subtracting "adjustment" persons with detailed characteristics. The number of people added or subtracted was determined by final adjustment factor for the post-stratum that the block part was in. If someone had to be added, the information from someone else in the block part who was counted in the census was duplicated. If someone had to be subtracted, the information for someone in the block part who was counted in the census was deleted.

Evaluations

The PES and adjustment process are based on many assumptions and have the potential for error. To evaluate the assumptions and potential error, the Census Bureau conducted numerous studies called P-studies because they referred to the PES. The studies were associated with the following general areas.

- Missing data on the PES questionnaire
- Misreporting of census day address on the PES questionnaire
- Fabrication of data in the PES by interviewers
- Errors in matching
- Errors in determining erroneous enumerations
- Balancing omissions with erroneous enumerations
- Correlation Bias (the tendency of the DSE to underestimate total population because some people are missed in both the PES and the Census)
- The homogeneity assumption

The results of these evaluations are essential to determining whether adjusted or unadjusted census counts are more accurate.

DEMOGRAPHIC ANALYSIS

The Census Bureau's other coverage measurement program was demographic analysis (DA). DA uses historical data on births, deaths, and legal immigration; estimates of emigration and undocumented immigration; and medicare data to develop an independent estimate of the population. The DA estimate of population is compared with the census count to yield another measure of net census coverage. DA can be only used to make reliable estimates at the national level. The DA coverage estimates were compared to the post-enumeration survey coverage estimates to assess the overall consistency of the two sets of estimates at the national level.

Birth and death records are available for the entire United States from 1933 on, but are not complete for years before 1933. Therefore, the Census Bureau had to find other ways to estimate the number of people who were born or died prior to 1933. In estimating births for each year, The Census Bureau added to the number of registered births an estimate of under-registration. Under-registration was estimated based on tests conducted in 1940, 1950, and 1964-1968. If the estimates of under-registration are off, they could have a significant effect on undercount estimates because birth data are by far the largest component in estimating the population through demographic analysis. Since national birth and death records are not available before 1933, the Census Bureau had to find other ways to estimate the size of the population 55 and older. For the population 65 and older, medicare estimates are used. For the population 55 to 64, estimates are made from revisions to earlier estimates.

The United States does not keep emigration records. Therefore, an estimate had to be made of persons who have left the country. While the United States does have good records of legal immigration, there is no accurate estimate of illegal immigration. The Immigration and Naturalization Service now collects different information than it did prior to 1980. That change further complicated the effort to estimate legal immigration. Also recent legislative reform allowing amnesty also complicated the issue since the Census Bureau did not know whether all of those obtaining amnesty actually reside in the United States. The Bureau used professional judgment to estimate the components of illegal immigration.

It is important to emphasize that results of demographic analysis are not exact but are estimates. To a large extent, they were based on assumptions and best professional judgment. As in the PES, the Bureau tried to estimate potential error in the data produced by demographic analysis in a series of studies call D-studies. Based on these studies, the Census Bureau developed a range of error around the demographic analysis estimates.

UNDERCOUNT STEERING COMMITTEE

To address the evaluation of the coverage in the census and the methods used to evaluate that coverage (the PES and DA), the Census Bureau formed the Undercount Steering Committee (USC). Their work was an important part of the July 1991 decision whether to adjust the full 1990 census for all uses. The work of the USC was also the major basis for the work done by CAPE. For a detailed description of the findings of USC, see Technical Assessment of the Accuracy of Unadjusted versus Adjusted 1990 Census Counts: Report of the Undercount Steering Committee, June 21, 1991.

BACKGROUND ON INTERCENSAL ESTIMATES

When the Secretary of Commerce announced his decision on July 15, 1991, not to adjust the 1990 census, he indicated his concern about the differential undercount. Because of that concern, he instructed the Census Bureau to continue its research into the area of potential adjustment. If the Census Bureau was able to resolve the technical problems associated with adjustment that were identified in the spring of 1991, then the Secretary asked the Census Bureau to consider incorporating results from the PES into the intercensal estimates program.

Basically, intercensal estimates are made by updating the most recent census base with estimates of population change (births, deaths, and net migration). Of course, the actual procedure is much more complicated and sophisticated. The Census Bureau makes estimates at the national, state, and county level every year and at the incorporated place (city) level every other year. These estimates have a variety of uses. Most notably, the estimates are used in funding allocations, as sample survey controls, and as denominators for many important statistics.

About one-third of the Federal funding programs use intercensal estimates of population as part of their funding formula, rather than using the 1990 census count for ten years. There may be items other than total population in the formula as well. The General Accounting Office has estimated that about 10 billion federal dollars a year are allocated based on funding formulas that use intercensal estimates[1]. States have within state fund-allocation programs as well. Many states use intercensal estimates to allocate within-state funding dollars.

Many sample surveys use national, and to some extent state, intercensal estimates as controls. The most notable is the monthly unemployment survey (the Current Population Survey, or CPS). Sample surveys generally have poorer coverage than a census; therefore, in order to improve the accuracy of estimates from a sample survey, the sample survey estimates are often controlled to an independent total (in this case, the intercensal estimate).

Many Federal agencies produce statistics per 1,000 persons (or some other base). Examples are crime statistics, incidence of certain health conditions, etc. The numerator of these statistics can be obtained at various points in time throughout the decade. In the absence of any updated information, calculating these kinds of statistics on a static 1990 denominator would be misleading; therefore, these Federal agencies use intercensal estimates of population as the denominator.

In order to be responsive to the Secretary's request on intercensal estimates, the Census Bureau formed the Committee to address the technical issues related to a potential adjustment of the base for intercensal estimates. The Committee was made up of many people who also served on the Undercount Steering Committee for the July 1991 decision. However, the Committee also

---

[1]Federal Formula Programs - Outdated Population Data Used to Allocate Most Funds (GAO/HRD-90-145, September 1991).

included some new members, including some Census Bureau staff very familiar with intercensal estimates. Though the Committee focused on the technical issues surrounding a potential adjustment, early in the Committee's deliberations, the Committee also had to make some key decisions related to the unique nature of the intercensal estimates program. The Committee decided that:

1. For the purpose of survey controls, there would be only one decision point in the decade about whether to adjust intercensal estimates.

2. If there was a decision to adjust, there would have to be a mechanism to make the intercensal estimates additive from the smallest area to the national total.

3. There would not be adjustment for some uses of intercensal estimates, but no adjustment for other uses of the estimates.

4. If there were a decision to adjust, the amount of the adjustment would be calculated on the base population. This adjustment plus an estimate of population change for the time period since the census would be added to the unadjusted base.

After every census, there is a change in the base used to calculate the intercensal estimates. Apart from the question of adjustment, there would be a change from a 1980 census base to a 1990 census base. For the use of estimates as survey control totals, that changeover date was postponed from January 1992 to January 1993. Therefore, 1992 estimates released in January 1993 would reflect the 1990 base. The postponement was made so that the decision on whether to adjust the base for intercensal estimates could be made at the same time. If there is a decision to adjust, then the change to a 1990 base and the change to a 1990 adjusted base would be simultaneous. If the decision is not to adjust, then there will be a change to the 1990 unadjusted base. In that case, even if evidence later in the decade would lead one to support adjustment, the base would not be changed from 1990 unadjusted to 1990 adjusted at a later point in the decade for the purpose of survey controls. Any change in base presents a discontinuity in uses based on intercensal estimates. Federal agency users of intercensal estimates for survey controls were quite clear that they strongly preferred only one such discontinuity during the decade.

On a technical basis, it is conceivable to be able to support adjustment at one level (say states), but not at lower levels. In such a case, state estimates would add to the national estimate, but substate estimates would not add to state estimates. There was agreement from users and from the staff making the estimates that failure to have additivity was not only undesirable, but close to unacceptable. Also, on a technical basis, it is conceivable to be able to support adjustment for one purpose (for example, national survey controls), but not for another (for example, subnational fund allocation). The Committee found this situation undesirable. Finally, it is possible for the Census Bureau to decide not to adjust the base of estimates but for some Federal agencies to do their own adjustment. This topic was discussed among Federal agencies at a meeting at the OMB. There was general agreement that it would be unacceptable to have variable sets of intercensal estimates used differently by different Federal agencies.

Estimates start with a base population and add estimated population change (births, deaths, and net migration). If estimates are adjusted, an additional term would be added that represents the net adjustment level for each area. This net adjustment level is the difference between the adjusted base population and the unadjusted base population. In the estimation process, the sum of this net adjustment and the estimated population change would be added to the unadjusted population base. Under this procedure, the net adjustment would remain constant throughout the decade.

FURTHER RESEARCH
THE BASIS FOR THE ASSESSMENT

When discussing the issue of whether to adjust the 1990 census, almost all experts agreed that with more time, there would be refinements and changes to the estimated undercount. Most experts, however, assumed these changes would be relatively small. Since the July 1991 decision, the Census Bureau had the time and at the direction of the Secretary of Commerce, continued to examine the estimated undercount. As expected, the Census Bureau has made some refinements and changes. During this analysis, the Census Bureau discovered a significant computer processing error in the system used to determine the undercount estimates that were under consideration in spring 1991. As a result of an error in computer processing, the estimated national undercount rate of 2.1% was overstated by 0.4%. After correcting the computer error, the national level undercount was estimated to be about 1.7%. After making other refinements and corrections, the national undercount is now estimated to be about 1.6%. Attachment 3 shows revised undercount estimates by selected age-sex-race categories. Attachment 4 shows revised undercount estimates by state. Attachment 11 shows revised undercount estimates for cities of 100,000 or more population. Attachment 12 shows revised undercount estimates for counties of 100,000 or more population.

Since PES undercount estimates were based on a sample survey, they are subject to error. There is sampling error to reflect the fact that the information came from some and not all of the population. The estimates are also subject to biases. For example, errors in matching, erroneous responses from respondents, etc. can bias the undercount estimate. Just as for the estimate of undercount, the Census Bureau also refined its estimates of bias. The level of total bias, excluding correlation bias[2], on the revised estimate of undercount is negative 0.73 (-0.73%). Therefore, about 45% (0.73/1.58) of the revised estimated undercount is actually measured bias and not measured undercount. In 7 of the 10 evaluation strata[3], 50% or more of the estimated undercount is bias. When correlation bias is included, these percentages go down. With correlation bias, the revised estimate of total bias is negative 0.35 percent (-0.35%). Including correlation bias, about 22% of the revised estimate of undercount is actually bias and not measured undercount. In general, the Committee was concerned that the estimate of correlation bias could be an underestimate, which meant the total bias estimate of negative 0.35% was an overstatement. There was limited time and methodology to investigate this concern further. The Committee did not feel lack of more information on this concern had an appreciable effect on their overall conclusion.

---

[2]Correlation bias is a term that reflects the fact that the DSE of total population based on the PES is an underestimate for the model used by the Census Bureau. The DSE is downwardly biased because of correlation bias which occurs, for example, because there are people missed in both the census and the PES. Correlation bias is described more fully below in the section entitled Third Issue-Part B, p 21.

[3]See Attachment 6 for a description of evaluation post-strata.

When the Committee began discussing the issue of whether to adjust the base for intercensal estimates, it started by reviewing the technical concerns raised about whether to adjust the 1990 census. This analysis produced a list of concerns, which the Committee summarized into five key areas.

1. Could the problems in the smoothing model, including lack of robustness, be resolved?

2. Could the estimated biases in the PES estimate of undercount be removed?

3. Were all components of the bias adequately reflected in the total error model, and was total error being accurately handled in loss function analysis?

4. Could we learn more about whether or not our homogeneity assumption held sufficiently to support adjustment?

5. Could we resolve the inconsistencies between the PES and other estimates of undercount, primarily Demographic Analysis?

There were other issues raised. While it would have been helpful to research these other questions as well, the Committee felt comfortable in confining its research efforts to the five key questions. The Committee felt they could make a reasoned choice about whether to adjust the base for intercensal estimates if they got appropriate information on these five issues.

FIRST ISSUE: COULD PROBLEMS IN THE SMOOTHING MODEL BE RESOLVED?

Summary: The Committee was very comfortable with the new post-stratification scheme which reduced sampling variance enough to avoid the use of smoothing. However, because of the limitations of artificial population analysis[4], there was still some concern with the finding that there was no loss in homogeneity[5] in a smaller post-stratum design that had only about 25% as many post-strata. (See fourth issue.)

For the July 1991 decision on whether to adjust the 1990 census, the sample of about 400,000 people was post-stratified into 1,392 groups. A person could be in one and only one of the 1,392 post-stratum groupihgs. Some of

---

[4]Artificial Population Analysis refers to the study to examine if the persons within each of the 357 post-strata were alike (homogeneous) with regard to their probability of being counted in the census. Artificial Population Analysis is described below in the section entitled Forth Issue, p 25.

[5]To make estimates from the PES, each sample person is assigned to one and only one post-stratum. A necessary assumption is that every person within a post-stratum has approximately the same chance of being counted in the census or the PES. This assumption is called the homogeneity assumption.

those post-stratum groupings were quite small so the estimate of undercount was subject to very high sampling variance. In order to reduce this sampling error, the Census Bureau used a technique called smoothing. Smoothing was a regression prediction model. Based on items correlated with undercount, the undercount for each of the 1,392 post-strata was predicted using the regression model. Then, the final undercount was an average of the predicted undercount and the directly observed undercount.

The smoothing process was successful at reducing the sampling variance. However, there were several issues raised about the entire smoothing process. It would have taken a large, intense, and uncertain research program to have answered all of these concerns. Therefore, the Committee chose a different approach. The Committee agreed to reduce the number of post-strata. By doing so, each new post-stratum would have more sample size than under the 1,392 system, and presumably, enough sample size so that the estimates would be stable (meaning the estimates would not have very large sampling variance); therefore, no smoothing would be required. It was expected that there would be some loss of homogeneity by going to a smaller post-stratum design, since with fewer strata, each stratum now had more people. Therefore, one could expect that it was less likely that everyone within these larger strata had the same capture probability as in smaller strata. The Committee assumed that the loss in homogeneity would be smaller than the problems and potential error from smoothing. As it turned out, the Committee's assumption seemed to be correct.

Based on measures of census performance and general patterns of undercount, a new set of 357 strata were designed. The 357 strata were not a simple regrouping of the 1,392 strata. The 357 strata design included 51 main strata defined by geography, owner-renter, and race/Hispanic cross classified by 7 age groupings cross classified by male-female. Attachment 5 contains a description of the 357 post-stratum design. This 357 design turned out to be a very effective stratification, primarily because we were able to examine additional data before defining the strata. Perhaps the most important piece of information for this examination was the strong relationship of living in owner or renter housing units to undercount. Hence, owner-renter status is very prominent in the 357 design.

We prepared revised PES estimates of undercount based on the 357 design and analyzed sampling variance by post-stratum. The intent was to verify the assumption that the sampling variances under the smaller (357) design would be relatively stable. At the state level, the variances were at an acceptable level[6]. Attachment 10 contains revised estimates of undercount or overcount for the 51 main post-strata that were part of the 357 post-stratum design.

The Committee was also concerned with the potential loss of homogeneity with the smaller post-stratum design. Using artificial population analysis, the Committee examined the homogeneity of the 1,392 design compared to the 357 design. Artificial population analysis is described below in the section called Fourth Issue. Based on the artificial

[6] C.A.P.E. minutes 4-6-92, Attachment 3.

population analysis assuming no bias in the PES, the Committee found the homogeneity for the 1,392 design and the 357 design to be about the same[7]. This result at first seemed counter-intuitive since one would have expected some reduction in homogeneity. However, the result may be explained by the fact that the 357 design is much more effective than the 1,392 design (probably true since the 357 design was based on a careful review of auxiliary data), by limitations of the artificial population analysis, or by a combination of both those factors.

In summary, the Committee was very comfortable with the new stratification. In general, for state-level estimates, the Committee felt satisfied with the 357 design without smoothing versus the 1,392 design including smoothing. However, because of the limitations of artificial population analysis, there was still some concern with the finding of no loss in homogeneity by going to a smaller post-stratum design that had only about 25% as many post-strata.

SECOND ISSUE: CAN ESTIMATED BIASES BE REMOVED FROM PES ESTIMATES?

> Summary: One of the first steps in further analysis of the PES was to re-examine the 104 blocks which had the greatest effect on the undercount. Many of the blocks had such a significant effect, they could be considered outliers. As a result of the examination of 104 blocks[8], corrections to the Post Enumeration Survey (PES) undercount estimates and bias removal were conducted. The net result was to reduce the estimated national net undercount by 0.1%. During that analysis, the Census Bureau also found and corrected a computer error that had incorrectly overstated the 2.1% undercount reported in July 1991 by .4%. The July 1991 estimate of undercount was reduced by 0.4% because of the computer error and an additional 0.1% because of modifications and bias removal resulting in a revised July 1992 national PES estimate of undercount of about 1.6%. The Committee obviously was satisfied that the decision to do a review of 104 blocks led to the discovery of the computer processing error. The Committee was also confident that outlier blocks had been more appropriately handled. As for bias removal, the Committee had mixed feelings. They were pleased that the review of only 104 blocks had removed a relatively large amount of bias. But, a significant amount still remained. The Committee could find no reliable or expedient method to remove the balance of the bias from the PES estimates.

The PES estimates of undercount are subject to biases. The Census Bureau had many evaluation programs to try to measure the level of these biases. At the U.S. level for total population, the estimated bias was negative 0.73% (or negative 0.35% if correlation bias is included) on an estimated

[7]C.A.P.E. minutes 4-6-92 Attachment 5 and C.A.P.E. minutes 3-9-92 Attachment 1.

[8]Small blocks were often combined to form block clusters. This report uses blocks to refer to blocks and block clusters.

undercount of about 1.6%. If it was possible, it would be desirable to remove these biases before any potential adjustment since the PES estimate of undercount including the bias is an overstatement of the undercount the PES actually measured. At the U.S. level for total population, the bias could be removed. The Committee discussed the possibility of removing the bias at sub-national levels. The only alternative was a modeling approach. Considering the very small samples used to estimate the biases and the difficulties of modeling, the Committee was very reluctant to try to remove the bias by modeling. The Committee was concerned that more error would be introduced than the level of error we were trying to remove. A further complication was the concern that our estimate of correlation bias was conservative (see page 15).

As a partial solution to bias removal, the Committee recommended an examination of the blocks that had the potential to contribute the most to the PES estimate of undercount. If the bias could be removed from these blocks, the PES estimates would be improved. Of course, the results from this set of blocks could not be generalized to other blocks, so any solution would only be a partial removal of the bias. 104 blocks were included in the study. The study is referred to by various names since additional components to the study were added over time. This study was originally called OCR (Outlier Cluster Review) because of the intent to review the blocks that had outliers. When the study was expanded to a second purpose (removal of bias), the study was called Selective Cluster Review (SCR).

During the SCR, several types of problems were examined. The treatment of outliers was reexamined and corrected as necessary. Some blocks had unusual results and had very big effects on the estimated undercount, effects far larger than one block should be expected to have. These are called outliers. They are similar to unusual marks by judges in athletic competitions. For the July 1991 estimates of undercount, there was a method to defuse the effect of these outliers. Now, with more time, we were able to reexamine these outliers and to use better methods (when applicable) to dampen their effect.

In addition, during SCR, we looked for errors. An example is failure to search in the proper block. Searching for matching should have been done in the PES sample block as well as the ring of blocks surrounding the sample block. Generally, this was done. Sometimes errors were made and the matchers failed to look into the entire ring. Mistakes like these were corrected.

Matching, even in the proper set of blocks, is error prone. Errors in matching can lead to a bias in the PES estimates. During SCR, expert matchers tried to remove all matching error and therefore any bias in the PES estimate due to matching.

As a result of all aspects of SCR, the estimated national undercount was reduced by one-tenth of one percent (0.1%)[9]. The bias reduction only applied to the 104 blocks and could not be generalized to other blocks. The 104 blocks represent about 2% of the total sample while the 0.1% reduction on an estimated 0.7% total bias represents about a 14% reduction. Even though total bias could not be removed, these numbers show that the effort of redoing these 104 blocks was well worth it. The results of the SCR were also subtracted as appropriate from the total bias so that the resulting total bias only represents residual error for residual blocks (the total minus these 104 blocks).

During the SCR, Census Bureau staff discovered a computer processing error that affected the estimates of undercount released in July 1991. Codes that were attached to cases in clerical processing were incorrectly fed into the computer processing. Errors went in both directions (increasing and decreasing the estimated undercount), but the net result of the error was to reduce the estimated national undercount of 2.1% by 0.4%.

THIRD ISSUE: IS THE TOTAL ERROR MODEL COMPLETE?

Summary: With regard to total error, the Committee was completely satisfied that all components of bias were represented. The Committee was concerned about the accuracy of some of the estimates of bias and the high variance for some estimates of bias. The general conclusion was to use caution in evaluating the results of loss function analysis since the target numbers in that analysis were so dependent on the levels of estimated bias. The Committee felt that correlation bias should be a component of total error. However, there was concern about our method of estimating it and very serious concern about the method of allocating it to states, cities, etc. Since there did not appear to be methods or time to analyze this allocation issue further, the Committee requested that loss function analysis be done with and without correlation bias. There was a choice of various loss functions. Primarily, the Committee concentrated on loss functions that examined proportionate population shares and not population counts. In addition, in general, the Committee considered loss functions based on squared error not absolute error. Using hypothesis tests with 10% significance, loss function analysis excluding correlation bias does not support adjustment. Using hypothesis tests with 10% significance and including correlation bias, all but one of the loss function analyses favors adjustment at the state level when examining aggregate loss. The Committee tended to accept these findings keeping in mind the numerous caveats. As a result of some comments from the Panel of Experts, the Committee was concerned about whether the significance level they used for the hypothesis tests was appropriate.

[9]"Post Census Rematching for the Outlier Cluster Review," Howard Hogan, undated; C.A.P.E. minutes 6-11-92 Attachment 1,2; C.A.P.E. minutes 4-20-92 Attachment 2.

THIRD ISSUE-PART A: TOTAL ERROR

The third major concern was whether the total error model contained all components of error and whether the components of error were adequately measured. In terms of whether all components of error were considered, two new components were added-- error due to cases done very late in the regular census (called late-late returns) and treatment of out-of-scope cases. The Committee felt completely confident that all components of error had been listed and considered.

The Committee could come to no agreement about the adequacy of the level of error measured for each of these components. There were concerns that matching error was determined by a dependent study and not an independent study. There were concerns that evaluation interviews used to determine the quality of the PES were conducted in February 1991. ten months after the census. There was concern that the estimate of only 13 fabrications in a sample of 150,000 seemed low compared to reasonable expectations. The Committee strongly agreed that the evaluation sample sizes were too small. The sampling error on several of the estimates of bias was extremely high.

In summary, with regard to total error, the Committee was satisfied that all components of error were represented. The Committee was concerned about the accuracy and variance of the estimates of bias, but there was really nothing that could be done. The general conclusion was to use caution in evaluating the results of loss function analysis since the target numbers in that analysis were so dependent on the levels of estimated bias. Attachment 6 contains estimates of the bias.

THIRD ISSUE-PART B: CORRELATION BIAS

The Committee spent a good deal of time discussing one aspect of total bias--correlation bias[10]. The Dual System Estimate (DSE) of total population produced by comparing the PES and the census is a biased estimate. It is biased because of matching error, etc. These components of bias are described immediately above.

The DSE can also be biased by correlation bias which has multiple components. The first is that the DSE assumes that a person's participation in the PES is not affected by his or her participation in the census (the causal independence assumption). Failure of this assumption can cause a bias. Generally lack of independence is not considered to be a big problem since the PES is conducted almost 4 months after the census and because of other controls introduced into the PES system.

The second component of correlation bias occurs because of variable capture probabilities within a post-stratum. The DSE does not require that the census and the PES have the same probability of counting people (called capture probability). But, the DSE does assume that within a post-stratum,

[10]Sometimes, model bias is used synonymously with correlation bias. In this report, correlation bias will be used.

everyone in the PES (or everyone in the census) has approximately the same capture probability. So, for example, a white male renter age 30-49 in rural areas of Louisiana is assumed to be just as likely to be counted as a white male renter age 30-49 in rural Mississippi, etc. Generally, if people within a post-stratum have differing capture probabilities, then the DSE is downwardly biased. That means the DSE underestimates the total population and in most cases would underestimate the undercount.

As a special case of variable capture probabilities, assume within a post-stratum there is a set of people with zero probability of being captured. These are often called the impossible to count or people missed in both the census and the PES. They are another component of correlation bias.

There are no direct estimates of either of these components of correlation bias, but an estimate for the total of both combined is obtained by comparing PES estimates to Demographic Analysis (DA) estimates. To estimate the level of correlation bias, the assumption is that sex ratios as determined by DA are accurate. Then, since in general the DSE estimates of males are lower than the DA estimates of males, there is a calculation of how many males would have to be added to the DSE to make the PES sex ratio equal to the DA sex ratio. These added males are an estimate of the level of correlation bias in the PES.

Actually, after estimating the extent of correlation bias, it is not added to the DSE of total population (just as other estimates of bias are not subtracted). Rather, the estimate of correlation bias is added to the total error model and is used to determine target numbers for loss function analysis.

The Committee was concerned about the combination of the two components of correlation bias, but there did not appear to be any alternative. The Panel of Experts expressed the same sentiment. They agreed that they were uncomfortable with the combination, but there does not seem to be an easy alternative. The Committee also was concerned that the PES measures more females than DA so that this method of estimating correlation bias should have had the effect of estimating a true population (for loss function analysis target numbers) that was bigger than total population in DA. However, the sum of the target populations did not equal the sum of the PES estimate and the level of correlation bias that was estimated to be added, as it should have. There was no time to examine these concerns further. Finally, there was concern that the method used for comparing the DSE with bias to DA understated the estimate of people missed due to correlation bias.

Mostly, however, the Committee was concerned with the method of allocating the correlation bias. Basically, the estimated missing people due to all types of correlation bias (all males) are allocated back to each post stratum proportional to the estimate of the number of males in the fourth cell of the DSE for the post-stratum. Further modeling is used to allocate the total error down to sub post-stratum levels.

The fourth cell in the DSE is an estimate of the number of people missed in both the PES and the census, but it is a biased estimate because of correlation bias. It is not directly estimated, but an estimate can be obtained by subtraction. Some of the numbers used in the subtraction are sample estimates, therefore, they are subject to sampling variability. The fourth cell is expected to be the product of the true population times one minus the capture probability of the PES times one minus the capture probability for the census. In theory, this number cannot be negative. But, in practice, due to sample variability, matching error, etc., it can be estimated to be negative. When the estimate in the fourth cell is negative, no amount of the estimated people missed due to correlation bias is allocated to that post-stratum.

Both the Committee and the Panel of Experts were very concerned about the negative values in the fourth cell. The Panel of Experts suggested some methods to change the DSE process to avoid negative values. There was also considerable concern about using the fourth cell as the basis for allocation of the estimate of people missed due to correlation bias. In fact, other methods of allocation had been tried by the Census Bureau.

In summary, the Committee felt that correlation bias should be a component of total error. However, there was concern about our method of estimating it and very serious concern about the method of allocating it. Therefore, the Committee requested that loss function analysis be done with and without correlation bias. Each Committee member would then have to make some judgements about how to analyze the results.

THIRD ISSUE-PART C: LOSS FUNCTION ANALYSIS

Estimates of bias in the PES estimates of undercount are useful for interpreting the accuracy of the PES estimates. But, estimates of bias were also a key component in a summary analysis called loss function analysis. If truth were known, the census count and the adjusted base count could be compared to truth and an appropriate choice could be made. That of course is impossible. To approximate that comparison, the Census Bureau performed loss function analysis.

As a first step in loss function analysis, the true population is estimated. This estimate is called the target population. It is estimated by taking the PES estimate of population and modifying that estimate based on the estimates of error in the PES (the components of bias from the total error model). These estimates of bias are also subject to error, so you can't simply subtract bias from the PES estimate and assume that is the true population. A further complication is that estimates of bias are only available for 10 evaluation post-strata and target numbers are needed for every state, every county, every place, etc. A modeling system is used to allocate the bias from the 10 evaluation post-strata to sub-levels of geography. Once target numbers are calculated, there is a comparison to see whether census counts or adjusted counts are closer to the target numbers, which are assumed to be "truth." There is still an issue of what is the appropriate comparison between census, adjusted and target numbers. Should it be a simple difference? If so, how are pluses and minuses handled? Should it be the square of the differences, which avoids the problem of pluses and minuses but overemphasizes states (or other areas of interest) with big differences. Or should it be some kind of weighted squared difference to avoid the over-effect of big states but to still reflect some of the differences in state size?

The Committee could come to no consensus on these difficult questions. Therefore, the Committee ran a variety of loss functions. These were a combination of:

- Various methods of allocating the bias to target numbers

- With and without correlation bias

- Absolute and squared error as well as variations of those to take account of variation in state (or other area of interest) size.

Even with these various loss functions, there was still another important question. Do you only look at the aggregate loss over all areas of interest (example, all states), or do you look at individual losses? This question was discussed with the Panel of Experts. The Panel felt that a simple count of "winners" and "losers" was inappropriate. One suggestion was to use a Pitman nearness measure. Time prevented that kind of analysis. In the absence of this measure, the Committee continued its original intent to examine aggregate loss. The Panel supported analysis of aggregate loss. In doing aggregate loss analysis, the Committee heeded the advice of the Panel of Experts who strongly recommended that loss function analysis be viewed only as a tool and not an exact decision mechanism

In examining total loss over a set of areas (like all states), there was a question about whether the difference in aggregate loss between the census and adjusted base counts was a real difference or only due to random error. The Census Bureau had developed a statistical hypothesis test to try to answer that question. The Panel of Experts reviewed this work as well. In particular, the representative from Statistics Canada, who face the same problem, commented on the proposed hypothesis test. That expert warned that in effect we were not doing a standard hypothesis test, but rather we would be making a decision on which set of estimates to use based on the results of the test. If we continued with the standard test, we could be making mistakes about what level of significance to use. The most appropriate level might very well be larger than the 10% level of significance the Committee chose to use. Because of the lateness of the suggestion, time prevented us from completely examining the alternative hypothesis test approach. Hence, the Committee used, with caution, the significance level of standard hypothesis test results.

In summary, using hypothesis tests with 10% significance, loss function analysis excluding correlation bias does not support adjustment. Using hypothesis tests with 10% significance and including correlation bias, all but one of the loss function analyses favors adjustment at the state level when examining aggregate loss[11]. The Committee tended to accept these findings keeping in mind the numerous caveats mentioned above.

FOURTH ISSUE: DOES THE HOMOGENEITY ASSUMPTION HOLD?

Summary: Just as in July 1991, the results on whether the homogeneity assumption holds are inconclusive. The new research used to examine the homogeneity assumption (called artificial population analysis) indicates that the assumption does not hold when the bias in the estimate gets to be about 25% or higher. Since the bias in the Post Enumeration Survey (PES) estimate as currently measured is 22% to 45%, the Committee was concerned.

An integral part of the PES/DSE system is to assume that everyone within a post-stratum has approximately the same probability of being counted in the PES. This is often referred to as having the same "capture probability." As discussed in the part of the third issue having to do with correlation bias, failure of this assumption leads to a bias in the DSE. It is also important because of the way the sample is selected and used to make estimates for states, cities, etc. Very few political units, including states, have direct estimates from the PES. That is, the state (or city) was not defined as a universe, and then a sample drawn from it to represent it. Rather, the sample was drawn by region, type of area (large urban area, other urban, rural), race, etc. Therefore, a sample case in Tennessee (for example) also is used in the estimate of undercount for Florida, Georgia, etc. This approach assumes homogeneity. Recognizing the importance of this assumption, the Census Bureau designed a study (labeled P-12) to analyze whether the homogeneity assumption held. The results of P-12 were mixed or inconclusive.

Recognizing this, the Committee asked for more extensive research into the issue of homogeneity. The new research was called artificial population analysis. Basically, items felt to be correlated with undercount were selected. They were called surrogate variables. These items were then scaled to the level of the undercount. For example, the mail return rate of census questionnaires was one of these items. The mail return rate was about 65% while undercount was about 2%. The 65% was scaled to 2%. Then an area that had a mail return rate 5% greater than the national average, got a scaled mail return rate 5% above the national average.

We know mail return rates for every area in the country. Using the same process used to estimate DSE's we estimated this scaled mail return rate. In effect, the comparison of the estimated scaled mail return rate to the known scaled mail return rate substitutes for the comparison of estimated undercount with known undercount.

---

[11]Summaries of loss function analysis results can be found in the following C.A.P.E. minutes: C.A.P.E. minutes 5-4-92 Attachment 4; C.A.P.E. minutes 6-1-92 Attachments 9-11; C.A.P.E. minutes 6-9-92 Attachment 5; C.A.P.E. minutes 7-6-92 Attachments 2,3.

Various types of loss function analyses were used to compare the estimated scaled surrogate variables with the actual scaled surrogate variables. If the loss from the estimate was small you could assume that the post-stratification was good and the homogeneity assumption was holding. If the loss was large, there would be cause for concern. In addition, we could examine the number of places (states, cities, etc.) "improved" by adjustment. We could do this kind of analysis for surrogate variables since we know truth (the actual value of the surrogate variable).

Based on artificial population analysis, a first analysis showed similar homogeneity for the 1,392 design as well as the 357 design as well as for a design with only 2 strata. Further analysis showed two problems. One, the surrogate variables did not vary much by post-stratum. Since the assumption was that undercount did vary by post-stratum, there was concern about whether this set of surrogate variables was a good set. Another concern was that the analysis assumed no bias in the surrogate variable estimates and the PES estimates of undercount are biased. Therefore, there was an attempt to find additional surrogate variables as well as to introduce bias into the artificial population analysis. Artificial population analysis was rerun with various levels of constant bias added. The bias in the PES is not constant, but there was no adequate way to introduce variable bias into the artificial population analysis.

The original five surrogate variables were:

-Allocation Rate (The rate at which questions without answers on the census questionnaire had to be allocated a response)
-Percent of population covered by the mail census procedure
-Percent enumerated by mail (mail return rate)
-Substitution rate (The rate at which an entire person's census characteristics had to be created by a computer algorithm)
-Percent of housing units that were multi-unit

The three additional items were:

-Percent in poverty
-Percent unemployed
-A mobility statistic

For states and most large geographic areas, without any bias, artificial population analysis supported the homogeneity assumption assuming that the surrogate variables act like undercount. Once bias is introduced, however, the artificial population analysis shows less and less homogeneity. When bias is 25% of the estimate, the artificial population analysis indicates that there is serious concern that the homogeneity assumption does not hold. Currently, with correlation bias included, the bias in the PES estimate of undercount is 22%. Without correlation bias, the bias is 45% of the estimate. In summary, the Committee could only support the homogeneity assumption with some concern since the level of bias in the PES was close to the point where artificial population analysis shows the homogeneity assumption fails to hold.

FIFTH ISSUE: CAN THE INCONSISTENCY OF PES AND OTHER ESTIMATES BE EXPLAINED?

Summary: Even though there were some points of concern, the Committee is much more comfortable with the consistency of the revised Post Enumeration Survey (PES) estimates and Demographic Analysis (DA) than they were with the July 1991 PES estimates and DA. At the state level, the Committee generally felt the revised PES estimates met their face validity expectations with some individual state exceptions.

As part of the July 1991 decision whether to adjust the 1990 census, there were many concerns about the PES estimates compared to other estimates, mainly Demographic Analysis (DA). In particular, there was concern that the PES estimated a higher population than DA and the fact that the PES estimated about a million more woman than DA. In addition, PES estimates were compared to "best professional judgement" estimates, mainly to see if undercount was being measured by the PES in areas where undercount was expected. This check was called face validity. Face validity checks, though not rigorous, indicated some areas of concern in the PES estimates. For these reasons, the Committee requested additional research to try to investigate the apparent differences.

With regard to DA, the revised PES estimates are now much more consistent. Attachment 7 contains a table summarizing the comparisons. The PES estimate of total population was now lower than the DA estimate, a more expected outcome. The estimated undercount from the PES at the national level was 1.6% compared to an estimate of 1.8% from DA. The PES estimate of women remained higher than DA (an unexpected result), but the difference has been reduced from one million to about 400,000 and was within sampling error. As expected, the PES estimates for Blacks (and in particular, young Black males) were much lower than the DA estimates. This is a result of correlation bias. Even though expected, the Committee was concerned about this problem because there was no method to adequately add these people back into PES estimates.

With regard to face validity checks, there also was now more consistency. Almost all of the changes between the revised PES and the July 1991 PES estimates were in the direction expected by the Committee.

Since intercensal estimates of states are of such importance, the Committee asked for an analysis of revised PES state estimates compared with other information on states to see if there was consistency. Basically, there was consistency with a few exceptions. The exceptions were substantiated by an independent analysis done by one of the Panel of Experts. The Committee was concerned about these exceptions, therefore, they could only conclude that, on average, there would be an improvement using adjusted base counts for states.

In summary, even though there were some points of concern, the Committee was much more comfortable with the consistency of the revised PES estimates and DA than they were with the July 1991 PES estimates and DA. At the state level, the Committee generally felt the revised PES estimates met their face validity expectations with some exceptions.

THE DECISION PROCESS

The decision process that led to the assessment of the Committee contained many parts. By far, the largest part was the year of extensive research and discussion between the Committee and the statistical staff at the Census Bureau. That part of the decision process is summarized in this report and recorded in far more detail in the minutes of the Committee. The decision process culminated with three key discussions. These were a day long meeting with the Panel of Experts, a decision discussion meeting with the Committee, and a decision discussion meeting with the Executive Staff of the Census Bureau. This section of the report summarizes those three meetings.

MEETING WITH PANEL OF EXPERTS:

The Census Bureau wanted to have outside review of the additional research it had done since July 1991. The Census Bureau wanted to include some Panel members who had not been too involved in the July 1991 decision in order to get a fresh look. In addition, the Census Bureau considered the outside expert advice it obtained in conjunction with the July 1991 decision. The Panel of Experts was sent materials in advance. In addition, each member was asked to chose two of five key areas on which to concentrate his or her attention. They were, of course, free to comment on any other issue, and as expected, they did. The meeting with the Panel was held on July 14, 1992. In order to place this summary of the Panel meeting in proper context, it is important to understand that the agenda for the Panel was restricted to major problems and that the Census Bureau specifically requested critical review.

In summary, the Panel made comments on the following key points:

1. The Panel thought the additional research done by the Census Bureau was extremely thorough and useful. The Panel took the time to commend the Census Bureau for this effort. They felt this research took the Census Bureau a long way towards being able to adjust at some time, even if not fully at the present.

2. The Panel thought the Census Bureau should only adjust for the geographic areas for which it was comfortable supporting the decision on technical grounds. Even then, there were bound to be some areas that were adversely affected by an adjustment or no adjustment, even though most were improved. The Panel urged the Census Bureau to examine the exceptions and see if they were "seriously" hurt. If so, the Panel recommended the Census Bureau reconsider an adjustment, even if it was technically defensible on average. For areas below the level for which there is technical backing to support adjustment, the decision about whether to adjust was more of a policy issue. The Panel did point out that errors in estimates of population change from the census year to the year of interest could be large, and perhaps larger than errors from adjustment, particularly for small areas.

3. The Panel cautioned that many of the statistical analyses used by the Census Bureau (Loss Function, Total Error Model, etc.) were just tools and not exact decision mechanisms.

4. The Panel would have felt more comfortable if the bias could be removed from the PES estimates before their use in any potential adjustment. The Census Bureau agreed with the concern of the Panel but knew of no adequate methodology to remove the bias by state, city, etc.

In addition, the Panel expressed some concerns:

1. The Panel was quite concerned about the negative values in the fourth cell. The Panel suggested ways to alter the DSE process in order to avoid the negative values.

2. While the Panel recognized the need to do something about correlation bias, they also recognized the potential problems caused by the inability to estimate the components of the bias separately. The Panel was also concerned about the problems with the proposed allocation scheme.

3. The Panel cautioned against loss function analysis where winners and losers were tallied up. Instead, if the intent is to examine individual losses/gains, the Panel recommended a Pitman nearness measure be used.

4. The Panel cautioned against too much reliance on the significance level in the hypothesis test the Census Bureau was planning to use and urged the Census Bureau to consider the implications of the approach to hypothesis testing being studied by Statistics Canada.

5. The Panel cautioned that artificial population analysis, like the P-12 study, was inconclusive about whether the homogeneity assumption held.

6. Some Panel members expressed concern about the extensive use of synthetic estimation in the adjustment process. (Examples: allocating undercount estimates to areas below which there were direct estimates, allocating bias, etc.)

Attachment 8 contains more detail from the meeting with the Panel of Experts.

C.A.P.E. DECISION DISCUSSION

In July 22, 1992, the Committee met with the Director to discuss each member's opinion about the accuracy of adjusted base counts for use in intercensal estimates. Prior to the main part of the meeting, one of the Committee members made a suggestion based on some analysis he had performed. He recommended the Committee consider a composite (50-50) estimate which would be the simple average of the census count and the adjusted base. The reasoning for the suggestion was that we have two estimates of population, both with error. Despite massive research, it is still inconclusive about which is better overall, for all levels of geography. Therefore, an average of the two might make sense. There is precedent for this kind of averaging in other Census Bureau work. Despite the lateness of the suggestion, the Committee members were asked to comment on the new proposal.

To help in the overall discussion about whether to adjust the base for intercensal estimates, there was a list of key uses and issues of intercensal estimates. Committee members were asked to tie their opinions about potential improved accuracy to the uses of the estimates and geographic level. The list is shown in Attchment 9.

Each Committee member expressed his or her opinion about whether or not the base for intercensal estimates should be adjusted. Though not unanimous, most of the Committee members felt that adjustment of the base should be done at the national and state level. For national and state uses of intercensal estimates, most Committee members felt adjusting the base would make the eventual estimates better on average. There was considerable concern about the states for which it was uncertain whether adjustment would make an improvement. Below the state level, the Committee could not make a recommendation about improvement from adjustment and supported the census counts. In terms of the issue of differential undercount and perception of fairness, the Committee strongly felt that adjustment at the state and national level would satisfy that element. The Committee could come to no agreement on whether an adjustment to the base would improve overall accuracy (accuracy at all levels of geography).

In addition to those summary findings, some other points were raised. These included:

1. No matter what the decision, the Census Bureau needed to examine the existing intercensal estimate challenge system[12]. Regardless of the Census Bureau decision on adjusting the base, a political jurisdiction who feels it was harmed by the Census Bureau decision can and will challenge.

2. Could we adopt the system used in Australia and perhaps Canada? The census is not adjusted, but intercensal estimates are.

---

[12]Currently, there is a challenge system in place that allows jurisdictions to question their intercensal estimates. The evidence supplied by the jurisdiction is reviewed by Census Bureau staff. The staff selected are not involved in the intercensal estimate operations. If the challenge is accepted, the intercensal estimate is changed.

3. No matter what the decision on adjustment of the base for intercensal estimates, the reliance on the current DSE system should be examined. Some of the problems with it might never be solved. (See the final section of this report-FUTURE)

The meeting closed with a discussion of the 50-50 composite suggestion. Only a minority of the Committee favored the 50-50 composite as a first choice, although many of the Committee members thought the composite could be a possible acceptable alternative. During the discussion, several pros and cons of the suggestion were listed.

PROS:

1. It would produce estimates that are additive. A procedure following the Committee's general consensus of states and higher would not be additive.

2. It is a move in the right direction. (This can also be viewed as a con since it is only a partial correction, even at the national level.)

3. It dampens the effect of noise (bias, error, etc.) in the PES and census.

4. At the substate level, the composite is probably better than the full adjustment.

5. Even with an adjustment, there would still be a benefit for respondents to take the effort to be counted in the future, because any potential adjustment based on the 50-50 composite method would only be a partial correction.

6. Analysis done by one Committee member showed that hypothesis test results at the state level were much more favorable to the composite estimate than to the full adjustment, even without including correlation bias.

CONS:

1. It is not as good an estimate at the national level as at the adjusted base, but it is probably a better estimate than an estimate with a fully adjusted base for substate levels. Substate improvement is at the expense of state and national estimates.

2. The two estimates (the DSE and the census) are not independent.

3. It was too late to fully examine the technical merits of the composite.

4. It is only half a solution to differential undercount.

5. It looks like a compromise or even like a "cop-out."

6. Why 50-50? 60-40 or some other combination might be better, and there is no way to know.

EXECUTIVE STAFF DECISION DISCUSSION

Following the Committee discussion, the Executive Staff of the Census Bureau met to give their views. Basically, the Executive Staff concentrated on policy concerns since the Committee had discussed the technical issues. The Executive Staff did not make a recommendation on whether or not to adjust the base for intercensal estimates, but rather raised some issues. The following points were raised at the Executive Staff meeting:

1. It is very important to make sure that people understand that the decision on whether to adjust the base for intercensal estimates is different from the decision whether to adjust the full census. Even if there is a decision to adjust the base for intercensal estimates, there is no intention to adjust the 1990 census because research shows insufficient technical justification.

2. The Census Bureau shoulo do what it thinks it can support based on statistical science.

3. The Census Bureau should consider the advice of users, but should not be forced into a decision because of pressure from users.

4. The Census Bureau should consider the effect of the decision on the public and in particular on its respondents.

5. The 50-50 composite suggestion looks arbitrary.

6. The adjustment issue is so complex, there is probably no single intellectually coherent solution. Most likely, none of the available options is fully consistent with the current research. Also, no matter what the decision, some people will not be satisfied.

On balance, the Executive Staff felt very strongly that there should be technical support for the eventual decision. The Executive Staff recognized that many issues, some of them nontechnical, would need to be balanced in making the final choice. Even so, it is very important for the Census Bureau to be confident about the technical support for the decision it chooses. Not only would the Census Bureau have to defend any decision, but the professionalism of the agency can be questioned if the Census Bureau cannot stand behind its decision on statistical grounds.

FUTURE

Regardless of the choice about whether to adjust the base for intercensal estimates, there were several concerns about the future raised during the final discussions. Generally, it was felt that the problem of differential coverage will continue in the future. Therefore, there were strong recommendations that research in the area of differential undercount should continue as input into the design of the year 2000 census. In particular, the following points were made.

1. The Census Bureau should examine alternatives to the Dual System Estimation process used in 1990. Some of the problems of that approach may continue despite best efforts, meaning that a full adjustment based on such a system might never be possible.

2. Even though it might not be statistically efficient, coverage measurement surveys in the future should have samples and estimation systems that produce direct estimates for key political areas (like states).

3. The Committee process was very successful and could be a good model for the future. Examples of the benefits included sufficient time, timely senior staff input, clear goals, etc.

4. Any proposed undercount estimation/adjustment scheme must be simple. It must be simple enough so the technical aspects can be evaluated and it must be simple enough so it can be explained, even to those without extensive statistical knowledge.

5. Methods of incorporating coverage measurement into the census process should be examined.

6. A system that produces one set of counts rather than unadjusted and adjusted counts is definitely preferred.

Attachment 1: List of

COMMITTEE ON ADJUSTMENT OF POSTCENSAL ESTIMATES (CAPE)

MEMBERS

Dr. Barbara Everitt Bryant	Director
Mr. C. L. Kincannon	Deputy Director
Mr. William Butz	Associate Director
Mr. Charles Jones	Associate Director
Dr. Robert Tortora	Associate Director
Mr. Peter Bounpane	Assistant Director
Ms. Paula Schneider	Chief, Population Division
Mr. John Thompson	Chief, Decennial Statistical Studies Division
Dr. Robert Fay	Senior Mathematical Statistician
Dr. Howard Hogan	Statistical Research Division
Dr. John Long	Population Division
Dr. Mary Mulry	Decennial Statistical Studies Division
Dr. Gregory Robinson	Population Division
Mr. Michael Batutis	Population Division
Mr. David Whitford	Decennial Management Division

Attachment 2

LIST OF MEMBERS OF PANEL OF EXPERTS WHO ATTENDED THE MEETING WITH THE CENSUS BUREAU

Mr. Don Royce
Senior Methodologist
Statistics Canada Social Survey Methods Division

Mr. Wesley Schaible
Associate Commissioner
Office of Research and Evaluation
Bureau of Labor Statistics

Dr. Fritz Scheuren
Director, Statistics of Income Division
Internal Revenue Service

Dr. Bruce Spencer
Department Head
Statistics Department
Northwestern University

Dr. Theresa A. Sullivan
Chair and Professor for the Department of Sociology
University of Texas at Austin

Dr. James Trussell
Associate Dean of Woodrow Wilson School and Director of the Office of Research
Princeton University

Mr. Joseph Waksberg
Chairman of the Board
WESTAT

Dr. Tommy Wright
Research Staff Member
Oak Ridge National Laboratory

Dr. Donald Ylvisaker
Director for the Division of Statistics, Mathematics Department
University of California

Dr. Alan Zaslavsky
Assistant Professor
Statistics Department
Harvard University

ATTACHMENT 3A: PES ESTIMATES OF UNDERCOUNT BY RACE AND SEX
JULY, 1992

Table 1 Table of PES Estimates for Selected Race/Origin/Sex Groups

		JULY, 1991 Original PES		JANUARY, 1992 Revised PES		JULY, 1992 357 PES	
Race/Hispanic/Sex	Census	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
Total	248709873	253979141	472946.472	252959473	461310.829	252712821	489754.595
Male	121239418	124249093	245445.426	123648997	238663.637	123623143	273518.304
Female	127470455	129730048	246737.086	129310476	241383.831	129089678	254912.175
Black	29986060	31505838	95559.460	31295058	93635.743	31377094	167925.028
Male	14170151	14974382	49052.934	14857391	47952.832	14900868	82912.806
Female	15815909	16531456	52914.183	16437667	51898.230	16476225	96609.126
Non-Black	218723813	222473303	424675.175	221664415	414933.642	221335728	453076.281
Male	107069267	109274711	222153.799	108791606	216160.510	108722274	249791.220
Female	111654546	113198592	220800.163	112872809	216539.374	112613453	239423.186
Asian or Pacific Islander	7273662	7504906	36264.289	7485602	36157.768	7447371	102828.516
Male	3558038	3688436	19879.800	3674532	19946.424	3684895	60817.829
Female	3715624	3816470	18469.115	3811069	18435.209	3762476	57240.421
American Indian	1878285	1976890	21726.014	1970537	21588.870	2051976	26259.820
Male	926056	980874	11512.232	977738	1130[illegible].066	1020059	13248.050
Female	952229	996016	10612.782	992799	10467.531	1031917	13252.478
Hispanic	22354059	23590274	103458.969	23471101	102033.476	23521183	180090.423
Male	11388059	12086513	57498.441	12008888	56356.003	12052241	114778.144
Female	10966000	11503761	52275.143	11462214	52082.441	11468942	84750.443

Table 2 Table of Undercount Rates for Selected Race/Origin/Sex Groups

		Original PES		Revised PES		357 PES	
Race/Hispanic/Sex	Census	UC Rt	SE(UC Rt)	UC Rt	SE(UC Rt)	UC Rt	SE(UC Rt)
Total	248709873	2.075	0.182	1.680	0.179	1.584	0.191
Male	121239418	2.422	0.193	1.949	0.189	1.928	0.217
Female	127470455	1.742	0.187	1.423	0.184	1.254	0.195
Black	29986060	4.824	0.289	4.183	0.287	4.433	0.511
Male	14170151	5.371	0.310	4.626	0.308	4.904	0.529
Female	15815909	4.328	0.306	3.783	0.304	4.008	0.563
Non-Black	218723813	1.685	0.188	1.327	0.185	1.180	0.202
Male	107069267	2.018	0.199	1.583	0.196	1.520	0.226
Female	111654546	1.364	0.192	1.079	0.190	0.852	0.211
Asian or Pacific Islander	7273662	3.081	0.468	2.831	0.469	2.332	1.349
Male	3558038	3.535	0.520	3.170	0.526	3.443	1.594
Female	3715624	2.642	0.471	2.504	0.472	1.245	1.502
American Indian	1878285	4.988	1.044	4.682	1.044	4.520	1.222
Male	926056	5.589	1.089	5.286	1.095	5.183	1.231
Female	952229	4.396	1.019	4.086	1.011	3.864	1.235
Hispanic	22354059	5.240	0.416	4.759	0.414	4.962	0.728
Male	11388059	5.779	0.448	5.170	0.445	5.511	0.900
Female	10966000	4.675	0.433	4.329	0.435	4.385	0.707

Note: Due to the nature of the data used to compute these counts for the 357 poststrata PES design, the American Indian counts in both Table 1 and Table 2 above include Eskimos and Aleuts for the 357 PES. The census count used for this group was 1,959,234. T census counts used to compute the original PES counts and the revised PES counts are shown in the tables.

ATTACHMENT 3B: REVISED PES ESTIMATES OF UNDERCOUNT BY AGE-RACE-SEX
JULY, 1992

Table 1 PES Estimates for Selected Race/Origin/Sex Groups for the 0 to 17 Age Group (357 Poststrata PES Design)

JULY, 1992

Race/Origin/Sex Group	Census	357 PES Estimate	Std. Error	Undercount Rate	Standard Error
Total	63604432	65695382	191195.568	3.183	0.282
Male	32584278	33649795	97745.288	3.166	0.281
Female	31020154	32045587	93459.542	3.200	0.282
Black	9584415	10311019	95917.245	7.047	0.865
Male	4849497	5215800	48390.736	7.023	0.863
Female	4734918	5095218	47527.287	7.071	0.867
Non-Black	54020017	55384363	172047.616	2.463	0.303
Male	27734781	28433994	88325.776	2.459	0.303
Female	26285236	26950369	83724.989	2.468	0.303
Asian or Pacific Islander	2083387	2152880	46537.029	3.228	2.092
Male	1063264	1099038	23792.412	3.255	2.094
Female	1020123	1053842	22745.817	3.200	2.089
American Indian, Eskimo, or Aleut	696967	742996	12481.466	6.195	1.576
Male	354875	378205	6315.004	6.169	1.567
Female	342092	364791	6166.491	6.222	1.585
Hispanic	7757500	8164834	77292.661	4.989	0.899
Male	3971164	4179630	39551.088	4.988	0.899
Female	3786336	3985204	37742.086	4.990	0.900

Table 2 PES Estimates for Selected Race/Origin/Sex Groups for the 18 to 29 Age Group (357 Poststrata PES Design)

JULY, 1992

Race/Origin/Sex Group	Census	357 PES Estimate	Std. Error	Undercount Rate	Standard Error
Total	48050811	49530134	192936.681	2.987	0.378
Male	24312055	25105216	129869.843	3.159	0.501
Female	23738756	24424918	113605.768	2.809	0.452
Black	6419397	6727151	60784.870	4.575	0.862
Male	3110320	3225832	38478.198	3.581	1.150
Female	3309077	3501319	41388.086	5.491	1.117
Non-Black	41631414	42802983	174778.637	2.737	0.397
Male	21201735	21879384	121313.350	3.097	0.537
Female	20429679	20923599	102738.356	2.361	0.479
Asian or Pacific Islander	1581231	1686549	47226.618	6.245	2.625
Male	802067	893983	35821.446	10.282	3.595
Female	779164	792566	31415.861	1.691	3.897
American Indian, Eskimo, or Aleut	414071	441408	7298.043	6.193	1.551
Male	210263	224725	4083.000	6.435	1.700
Female	203808	216683	3782.708	5.942	1.642
Hispanic	5525130	5903999	83906.191	6.417	1.330
Male	2984897	3207779	67903.944	6.948	1.970
Female	2540233	2696220	31412.026	5.785	1.098

ATTACHMENT 3B: REVISED PES ESTIMATES OF UNDERCOUNT BY AGE-RACE-SEX
JULY, 1992

Table 3 PES Estimates for Selected Race/Origin/Sex Groups for the 30 to 49 Age Group
(357 Poststrata PES Design)

Race/Origin/Sex Group	Census	July, 1992 357 PES Estimate	Std. Error	Undercount Rate	Standard Error
Total	73314363	74327349	178380.748	1.363	0.237
Male	36281757	36965692	114336.225	1.850	0.304
Female	37032606	37361657	94874.030	0.881	0.252
Black	8300318	8705762	57437.333	4.657	0.629
Male	3841762	4099633	38014.164	6.290	0.869
Female	4458556	4606129	31219.727	3.204	0.656
Non-Black	65014045	65621588	168451.681	0.926	0.254
Male	32439995	32866059	106016.209	1.296	0.318
Female	32574050	32755528	90532.426	0.554	0.275
Asian or Pacific Islander	2373785	2396349	35297.064	0.942	1.459
Male	1128527	1127567	23875.089	-0.085	2.119
Female	1245258	1268782	19001.048	1.854	1.470
American Indian, Eskimo, or Aleut	543821	560400	5746.845	2.958	0.995
Male	263525	276134	2812.700	4.566	0.972
Female	280296	284266	3232.422	1.397	1.121
Hispanic	5961207	6271153	61500.742	4.942	0.932
Male	3029043	3225477	40130.964	6.090	1.168
Female	2932164	3045676	33430.513	3.727	1.057

Table 4 PES Estimates for Selected Race/Origin/Sex Groups for the 50 and Older Age Group
(357 Poststrata PES Design)

Race/Origin/Sex Group	Census	July, 1992 357 PES Estimate	Std. Error	Undercount Rate	Standard Error
Total	63740267	63159956	164191.819	-0.919	0.262
Male	28061328	27902440	91400.020	-0.569	0.329
Female	35678939	35257516	98575.330	-1.195	0.283
Black	5681930	5633162	34874.194	-0.866	0.624
Male	2368572	2359603	22227.003	-0.380	0.946
Female	3313358	3273559	19516.989	-1.216	0.603
Non-Black	58058337	57526794	159823.396	-0.924	0.280
Male	25692756	25542837	89232.535	-0.587	0.351
Female	32365581	31983957	96067.229	-1.193	0.304
Asian or Pacific Islander	1235259	1211593	20586.691	-1.953	1.732
Male	564180	564307	7192.919	0.023	1.274
Female	671079	647286	18017.833	-3.676	2.886
American Indian, Eskimo, or Aleut	304375	307172	3091.413	0.911	0.997
Male	138523	140996	1832.019	1.754	1.277
Female	165852	166176	1554.022	0.195	0.933
Hispanic	3110222	3181198	45726.253	2.231	1.405
Male	1402955	1439356	27996.289	2.529	1.896
Female	1707267	1741842	32679.612	1.985	1.839

ATTACHMENT 4:

JULY, 1992

State Level Estimates and Estimated Undercount Rates

State	1990 Census	Original PES July 1991 Estimate	Original PES July 1991 UC Rt	Original PES July 1991 SE(UCRt)	357 PES July 1992 Estimate	357 PES July 1992 UC Rt	357 PES July 1992 SE(UCRt)
01 Alabama	4040587	4146133	2.546	0.383	4113119	1.763	0.316
02 Alaska	550043	560727	1.905	0.437	561255	1.998	0.364
04 Arizona	3665228	3790186	3.297	0.466	3754297	2.373	0.455
05 Arkansas	2350725	2402925	2.172	0.417	2392291	1.738	0.337
06 California	29760021	30888075	3.652	0.420	30594537	2.728	0.379
08 Colorado	3294394	3376099	2.420	0.470	3363357	2.050	0.383
09 Connecticut	3287116	3305658	0.561	0.556	3308309	0.641	0.406
10 Delaware	666168	686661	2.984	0.437	678372	1.799	0.377
11 District of Columbia	606900	638747	4.986	0.517	628309	3.407	0.901
12 Florida	12937926	13277708	2.559	0.384	13196855	1.962	0.390
13 Georgia	6478216	6632561	2.327	0.368	6618829	2.124	0.331
15 Hawaii	1108229	1136417	2.480	0.537	1129162	1.854	0.808
16 Idaho	1006749	1035271	2.755	0.501	1029213	2.183	0.434
17 Illinois	11430602	11592305	1.395	0.352	11544433	0.986	0.358
18 Indiana	5544159	5585918	0.748	0.370	5572239	0.504	0.399
19 Iowa	2776755	2807238	1.086	0.455	2788378	0.417	0.404
20 Kansas	2477574	2506427	1.151	0.353	2494762	0.689	0.350
21 Kentucky	3685296	3767824	2.190	0.418	3745662	1.612	0.370
22 Louisiana	4219973	4332297	2.593	0.366	4313516	2.169	0.339
23 Maine	1227928	1240076	0.980	0.611	1237124	0.743	0.562
24 Maryland	4781468	4868990	1.798	0.444	4882324	2.066	0.418
25 Massachusetts	6016425	6039315	0.379	0.548	6045161	0.475	0.485
26 Michigan	9295297	9403964	1.156	0.368	9361331	0.705	0.371
27 Minnesota	4375099	4419180	0.998	0.355	4394680	0.446	0.380
28 Mississippi	2573216	2632412	2.249	0.397	2628899	2.118	0.434
29 Missouri	5117073	5184411	1.299	0.352	5149052	0.621	0.363
30 Montana	799065	822092	2.801	0.514	818305	2.351	0.492
31 Nebraska	1578385	1594894	1.035	0.380	1588698	0.649	0.366
32 Nevada	1201833	1231620	2.419	0.469	1230675	2.344	0.383
33 New Hampshire	1109252	1115972	0.602	0.530	1118610	0.837	0.546
34 New Jersey	7730188	7836174	1.353	0.498	7774411	0.569	0.612
35 New Mexico	1515069	1586489	4.502	0.514	1563123	3.074	0.505
36 New York	17990455	18304414	1.715	0.451	18261955	1.487	0.581
37 North Carolina	6628637	6814693	2.730	0.363	6753175	1.844	0.347
38 North Dakota	638800	647837	1.395	0.463	643042	0.660	0.502
39 Ohio	10847115	10933439	0.790	0.354	10921925	0.685	0.360
40 Oklahoma	3145585	3213646	2.118	0.386	3202730	1.784	0.338
41 Oregon	2842321	2898058	1.923	0.445	2896147	1.859	0.401
42 Pennsylvania	11881643	11956891	0.629	0.477	11916630	0.294	0.483
44 Rhode Island	1003464	1006150	0.267	0.556	1004811	0.134	0.590
45 South Carolina	3486703	3589808	2.872	0.407	3558918	2.029	0.362
46 South Dakota	696004	706954	1.549	0.494	702878	0.978	0.548
47 Tennessee	4877185	5012173	2.693	0.386	4963686	1.743	0.344
48 Texas	16986510	17550747	3.215	0.378	17469248	2.763	0.395
49 Utah	1722850	1757423	1.967	0.537	1753121	1.727	0.497
50 Vermont	562758	570651	1.383	0.709	569091	1.113	0.765
51 Virginia	6187358	6352705	2.603	0.351	6313620	2.000	0.353
53 Washington	4866692	4986607	2.405	0.433	4957987	1.841	0.437
54 West Virginia	1793477	1842267	2.648	0.436	1819004	1.403	0.430
55 Wisconsin	4891769	4923844	0.651	0.369	4921997	0.614	0.397
56 Wyoming	453588	466067	2.678	0.481	463569	2.153	0.416
United States Totals	248709873	253979140	2.075	0.182	252712822	1.584	0.191

UC Rt — Undercount Rate as estimated from the PES.

SE(UCRt) — The sampling error of the estimated undercount rate.

ATTACHMENT 5: THE 357 POSTSTRATUM DESIGN FOR POSTCENSAL ESTIMATION--JULY, 1992

The following page defines the 51 poststrata groups and seven age sex groups used to poststratify the Post-Enumeration Survey (PES). These were used to develop dual system estimates for use in the postcensal estimation program. Cross classification of the 51 poststrata groups with the seven age sex groups yields 357 poststrata cells for which dual system estimates have been developed.

The following rough definitions are used:

"Urbanized area 250,000+" means that the PES sample block was part of an Urbanized Area the total population size of which was greater than 250,000.

"Other-urban" refers to all PES blocks that were part of an Urbanized Area not greater than 250,000 or were part of an other urban place.

"Non-urban" means all rural areas and other areas not falling into the above categories.

"Owner/Non-Owner" is determined from the tenure variable on the PES questionnaire. All persons in group quarters are non-owners by definition.

"Asian and Pacific Islander" refers to all people who report themselves as being Asian and Pacific Islander. This group is not restricted to the West or Mid Atlantic as it was in the July, 1991 estimates. Asians and Pacific Islanders of Hispanic origin are included here.

"American Indians on Reservations" include American Indians living on reservations and Tribal Trust Lands. All other concepts (Black, Non-black Hispanic, etc.) are defined as in the census.

"North East" states are Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont, New Jersey, New York, and Pennsylvania.

"South" states include Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia, Alabama, Kentucky, Mississippi, Tennessee, Arkansas, Louisiana, Oklahoma, and Texas.

"Midwest" states are Illinois, Indiana, Michigan, Ohio, Wisconsin, Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota.

"West" states include Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming, Alaska, California, Hawaii, Oregon, and Washington.

Revised Poststratification for Postcensal Estimation (357 Design) - 1992

	North East	South	Mid West	West
Non-Hispanic White & Other				
Owner				
Urbanized Areas 250,000 +	1	2	3	4
Other Urban	5	6	7	8
Non-Urban	9	10	11	12
Renter				
Urbanized Areas 250,000 +	13	14	15	16
Other Urban	17	18	19	20
Non-Urban	21	22	23	24
Black				
Owner				
Urbanized Areas 250,000 +	25	26	27	28
Other Urban	29*			
Non-Urban	30*			
Renter				
Urbanized Areas 250,000 +	31	32	33	34
Other Urban	35*			
Non-Urban	36*			
Non-Black Hispanic				
Owner				
Urbanized Areas 250,000 +	37	38	39	40
Other Urban	41*			
Non-Urban	42*			
Renter				
Urbanized Areas 250,000 +	43	44	45	46
Other Urban	47*			
Non-Urban	48*			
Asian & Pacific Islander				
Owner	49*			
Renter	50*			
American Indians on Reservations	51*			

* Indicates that the group is combined across all regions.

Age-Sex Groups

	Males	Females
0 to 17	a	
18 to 29	b	e
30 to 49	c	f
50 and Over	d	g

ATTACHMENT 6:

Total Error of the Net Undercount Rate Assuming No Correlation Bias and Synthetic Estimation of Net Component Errors

JULY, 1992

Evaluation Poststratum	$\hat{U}$*	$\hat{B}(\hat{U})$	St. Dev. $\hat{B}(\hat{U})$	Total St. Dev.	95% Interval
Non-Hispanic White and Other, Owner					
Urban 250k+	-0.50	0.32	0.99	1.06	(-2.95, 1.31)
Other Urban	0.11	0.21	0.25	0.34	(-0.79, 0.59)
Non-Urban	-0.22	0.86	0.87	1.00	(-3.07, 0.92)
Non-Hispanic White and Other, Non-Owner					
Urban 250k+	2.33	-0.06	0.60	0.96	(0.47, 4.32)
Other Urban	2.92	1.70	0.82	1.13	(-1.03, 3.47)
Non-Urban	5.30	0.47	0.74	1.35	(2.13, 7.53)
Black, Non-Black Hispanic, Asian and Pacific Islander, Urban 250k+					
Owner	1.33	0.84	0.44	0.67	(-0.86, 1.82)
Non-Owner	7.13	0.80	0.48	0.94	(4.44, 8.21)
Black, Non-Black Hispanic, Asian and Pacific Islander, Other Urban & Non-Urban					
Owner	2.07	2.38	0.90	1.25	(-2.81, 2.18)
Non-Owner	6.44	3.98	0.94	1.63	(-0.80, 5.72)
National	1.61	0.73	0.30	0.36	(0.17, 1.60)

*Based on PES population only.

Total Error of the Net Undercount Rate Assuming Synthetic Estimation of Net Component Errors

Evaluation Poststratum	$\hat{U}$*	$\hat{B}(\hat{U})$	St. Dev. $\hat{B}(\hat{U})$	Total St. Dev.	95% Interval
Non-Hispanic White and Other, Owner					
Urban 250k+	-0.50	0.31	0.99	1.06	(-2.94, 1.32)
Other Urban	0.11	0.18	0.25	0.34	(-0.76, 0.62)
Non-Urban	-0.22	0.81	0.88	1.00	(-3.03, 0.97)
Non-Hispanic White and Other, Non-Owner					
Urban 250k+	2.33	-0.68	0.76	1.07	(0.87, 5.16)
Other Urban	2.92	1.54	0.84	1.14	(-0.90, 3.65)
Non-Urban	5.30	-0.12	0.90	1.45	(2.52, 8.31)
Black, Non-Black Hispanic, Asian and Pacific Islander, Urban 250k+					
Owner	1.33	0.80	0.45	0.68	(-0.83, 1.87)
Non-Owner	7.13	-1.37	1.30	1.54	(5.42, 11.56)
Black, Non-Black Hispanic, Asian and Pacific Islander, Other Urban & Non-Urban					
Owner	2.07	2.23	0.95	1.28	(-2.71, 2.41)
Non-Owner	6.44	3.55	1.05	1.70	(-0.50, 6.28)
National	1.61	0.35	0.33	0.38	(0.50, 2.03)

*Based on PES population only.

ATTACHMENT 7: COMPARISON OF REVISED PES ESTIMATES VERSUS DA--JULY 1992

Comparison of the Census, Post Enumeration Survey (PES) and Demographic Analysis (DA) Estimates of Population and Percent Net Undercount: 1990

(A positive difference means that the demographic estimate is higher than the PES estimate; a negative difference means that the demographic estimate is lower).

Race, Sex, Age	PES Net Undercount				DA Net Undercount		Difference in DA and PES Percent Net Undercount	
	Original Estimates July 1991		Revised Estimates July 1992					
	Amount (1)	Percent (2)	Amount (3)	Percent (4)	Amount (5)	Percent (6)	Original PES 7=6-2	357 PES 8=6-4
TOTAL	5,269,267	2.07	4,002,947	1.58	4,683,913	1.85	-0.23	0.26
Male	3,009,674	2.42	2,383,724	1.93	3,480,216	2.79	0.37	0.86
Female	2,259,593	1.74	1,619,223	1.25	1,203,697	0.94	-0.81	-0.32
BLACK	1,519,776	4.82	1,391,033	4.43	1,836,272	5.68	0.86	1.25
Male	804,233	5.37	730,717	4.90	1,338,380	8.49	3.12	3.59
Female	715,543	4.33	660,316	4.01	497,892	3.01	-1.32	-1.00
NONBLACK	3,749,491	1.69	2,611,914	1.18	2,847,641	1.29	-0.40	0.11
Male	2,205,441	2.02	1,653,007	1.52	2,141,836	1.97	-0.05	0.45
Female	1,544,050	1.36	958,907	0.85	705,805	0.63	-0.73	-0.22

NOTE: Original PES estimates are the July 15, 1991 estimates based on 1392 poststrata and incorporate smoothing; revised PES estimates are the July 1992 estimates based on 357 poststrata, all PES revisions since July 1991, and no smoothing.

ATTACHMENT 8: THE MEETING WITH THE PANEL OF EXPERTS

While the Panel came to no consensus about whether the base for intercensal estimates should be adjusted, the Panel was extremely impressed with the extensive research done by the Census Bureau. The concerns raised by the Panel were not criticisms of the Census Bureau's work, but rather were indications of the difficulty and complexity of the overall issue as well as the fact that some of these problems may never be fully solved. The Panel concentrated its discussion on five areas as requested by the Census Bureau. These were the most difficult problem areas that Census Bureau statisticians had not been able to fully resolve. Not only was the discussion limited to difficult problem areas, but as requested by the Census Bureau, the Panel members were critical and raised concerns. Reading just a list of concerns can lead to an unbalanced view of what Panel members felt about the adjustment issue in general. Therefore, the parameters under which the Panel operated should be kept in mind in order to put the following more detailed discussion of Panel concerns in proper perspective.

FIRST AREA: TOTAL ERROR MODEL INCLUDING CORRELATION BIAS

During this discussion the Panel mentioned that it didn't see an easy alternative to the current method of treating correlation bias, but Panel members were uneasy about certain aspects of it. For one, the Panel was quite concerned about the negative fourth cells. In addition, there was concern that we weren't estimating the level of the bias properly. In particular, one Panel member felt we should consider comparing the unbiased PES estimates (taking out the bias) to DA in order to estimate the level of correlation bias. Another panel member expressed serious concern that the Census Bureau assumed all correlation bias was male. This panel member pointed to his research to show that there also are problems of differing capture probabilities in the female population. Currently, the Census Bureau's treatment of correlation bias assumes that doesn't occur. It was also during this discussion that most of the Panel recommended that the Census Bureau try to remove the bias from the PES estimates before making any adjustment. Another panel member went through the PES/DSE process in some detail with an emphasis on whether or not it was understandable to an average person and whether or not it was creditable. He pointed out several parts of the process that were of concern to him particularly the extensive use of synthetic estimation. He also cautioned that if new research between July 1991 and the present uncovered new findings, then he wouldn't be surprised to see additional research after July 1992 turn up new results and new estimates of undercount. Another Panel member strongly desired that total error be broken out separately by persons of Hispanic ethnicity. This section of the meeting concluded with a discussion of the problem of inconsistent race classification between systems (example: PES and DA), which the Panel felt was a significant issue that needed further research.

SECOND AREA: LOSS FUNCTION ANALYSIS

This part of the meeting was quite technical, with a review of the various loss functions under consideration. Most of the Panel advised against counting up winners and losers (For example: states that gained or lost in a loss function analysis done on states). Instead one Panel member recommended a Pitman nearness measure which he uses when faced with this kind of problem. Then, there was a discussion of aggregate loss. The Panel pointed out that decisions on aggregate loss may make sense statistically, but that the "losing" political areas might have a problem. Also, it was during this discussion that the Panel made a recommendation that the results of loss function analysis be used with caution. Loss function analysis is a tool, depends on personal standards of judgement, and is not an exact decision mechanism. It also was during this discussion that the Panel reiterated a theme they raised in the first topic. Panel members were concerned that there is too much confusion about the undercount/adjustment issue by the "person on the street." The Panel recommended that the Census Bureau try to alleviate that in the future. Finally, there was a discussion about the large number of states for which it doesn't matter much whether or not there is an adjustment. Both sides of the case were discussed. If so, why bother to adjust?; or if so, adjust all states in order to correct a problem in a few states and the error in most other states won't be too bad. This discussion ended with another theme heard often. The total error model is a good tool to try alternative assumptions. It is not an exact decision mechanism.

THIRD AREA: HYPOTHESIS TESTS

The Census Bureau had recognized the limitations of loss function analysis. In particular, once you had two losses to compare, was the difference between them a "real" difference, or could it be attributable solely to chance since these were sample estimates. To help answer that question, the Census Bureau planned some statistical hypothesis tests. The Panel was asked to review the Census Bureau plans.

This part of the discussion was led by the expert from Statistics Canada, since Statistics Canada was faced with a similar problem. The discussion was extremely technical. Before getting to the issue of the hypothesis test, the Panel member cautioned that several key questions had to be answered, and they all had an effect on the eventual hypothesis test. These questions included:

What is the quantity of interest? (Total population, population share, etc.)

Which Loss Function would be used?

How accurate are your target numbers?

How do you account for error in estimating the target numbers?

The bulk of the discussion centered about the technical performance of the hypothesis test assuming the above questions had been answered satisfactorily. Basically, the Panel pointed out that we were not simply dealing with a standard hypothesis test. Instead, we planned to use one of the set of estimates based on the results of the hypothesis test. Under those conditions, a model could be developed to examine the true level of risk for the hypothesis test. At present, Statistics Canada had developed such an approach. The Panel member urged the Census Bureau to take this finding into account in the significance level of the Census Bureau's proposed hypothesis test. During this part of the discussion, this panel member warned that if there is a high positive bias in the estimate of undercount, then the hypothesis test can be misleading, and in fact, adjustment can be very problematic when the estimate of undercount has a large bias. Also, it was pointed out that Statistics Canada feels its estimates of undercount at the province level are adequate for use in adjusting intercensal estimates, but not at sub-province level. Whether or not to adjust below the Province level will be more a policy call than a technical decision. Finally, it was during this part of the meeting that the Panel repeated its recommendation that if estimates of bias are good enough for use in determining target numbers for loss function analysis, then they should be removed from the PES estimates before any potential adjustment.

FOURTH AREA: ARTIFICIAL POPULATION ANALYSIS

Because of the way the PES/DSE system operates, the homogeneity assumption is a key one. In conjunction with the July 1991 decision, the Census Bureau studied homogeneity and recorded the results in study called P-12. Since the homogeneity assumption was so key, the Census Bureau undertook additional work in a study called Artificial Population Analysis. The Panel was asked to examine various aspects of the analysis. The Panel member who did part of the P-12 study led the discussion. The Panel member started with a brief review of study P-12 which he characterized as inconclusive. In reviewing the artificial population analysis, he thought the Census Bureau had taken a major additional step to try to investigate the issue, but he still felt the results were inconclusive. In his opinion, only two of the eight surrogate variables considered by the Census Bureau were associated enough with undercount to be considered. (Percent enumerated by mail and substitution rate.) He wondered if there were better alternative surrogate variables. The Panel also expressed some concern about the constant scaling of the surrogate variables to undercount. Variable scaling might be preferred. Likewise, the Panel was concerned about the constant introduction of bias into the artificial population analysis. Once again, variable bias would be preferred. Even so, the Panel was concerned that artificial population analysis showed failure of the homogeneity assumption when the constant bias was 25% or greater. One panel member did some work on his own. From that study, he concluded that by using substitution rate, adjustment looks better. Using poverty, the results are mixed. And, using unemployment rate, the census looks better. This kind of analysis supports the conclusion that even with all the new research, the results are inconclusive. This panel member felt that a considerable amount of additional work would be needed to get a definitive answer on whether the homogeneity assumption held.

FIFTH AREA: COMPARISON OF PES TO DA

Generally, at the national level, estimates of population from DA are felt to be "better" than estimates from a post-censal survey. Even so, the DA estimates are subject to some error. Before discussing the comparison of the PES and DA, one panel member shared her work on the quality of DA numbers. In addition to the known problems with DA, she pointed out some additional places where the DA estimates could be in error. These included:

1. Over correction for the under-registration of black males. (This error has the effect of overestimating the undercount.)

2. The problem of Mexicans near the border who register the birth in the US, but then return to Mexico to raise the child. (This problem has the effect of overstating the undercount.)

3. Under reporting of infant deaths near the border since the birth certificate can be resold. (This problem overstates the undercount.)

4. Concerns about the consistency and reliability of reporting data on vital statistics forms, especially those done by a third party. (These types of errors might not effect the estimate of total undercount, but would effect the estimates by age-race-sex.)

5. Concern about a change in a person's self perception of race/Hispanic over time. These characteristics could be recorded one way at birth and another at death. (This problem only has an effect on DA estimates of undercount by race/Hispanic.)

Even with these and other problems, there is still general confidence in the DA estimates, particularly at the national level. That is why the Panel was concerned about some inconsistencies between the PES and DA. In particular, one panel member reviewed the Census Bureau work that compared PES estimates by state with DA and other information. She was quite concerned about the states that seemed quite inconsistent. At this point, another panel member indicated that another independent study he had done confirmed the inconsistency in a similar set of states. The Panel discussed the issue and concluded that in an adjustment where there would be overall improvement for states, some states would be adversely affected, even if most were improved and the US average was improved. The Panel strongly recommended that the Census Bureau examine if these exception states were hurt "seriously."

The meeting closed with a brief discussion of the actual mechanism of the intercensal estimate process. During that discussion, there was a question about the accuracy of intercensal estimates. That question couldn't be answered exactly, but there was some summary information provided. Basically, by comparing the estimate in a census year to the census count, you can estimate the error in the estimates over a 10-year period. The following table summarizes the Census Bureau findings.

AREA	LEVEL OF ERROR OVER 10 YEARS[1]
States	1.5 - 2.5%
Places over 50,000	4.0%
Places 5,000 to 50,000	7.0 - 8.0%
Places under 5,000	16.0 - 20.0%

[1]Level of error as measured in previous decades. These error estimates exclude any estimated undercoverage in the census.

ATTACHMENT 9: USES OF INTERCENSAL ESTIMATES AND ISSUES CONSIDERED BY C.A.P.E

Uses of Intercensal Estimates:

1. Survey controls
2. Denominators for per capita Federal statistics
3. Funding programs
   a. State populations either for direct funding or as the first tier in a funding program
   b. Substate areas of 100,000 population or larger
   c. Substate areas below 100,000 population

Other Concerns:

1. National population estimates
2. Differential undercount and the perception of fairness
3. Overall accuracy

ATTACHMENT 10: ESTIMATED UNDERCOUNT/OVERCOUNT FOR 51 POST- STRATA, JULY 1992

Post- Strata Groups	Percent Undercount				Standard Errors			
	North East	South	Mid West	West	NE	S	MW	W
Non-Hispanic White & Other								
Owner								
Urbanized Areas 250,000+	-2.13	0.68	-0.26	0.34	1.08	0.71	0.39	0.65
Other Urban	-1.08	0.52	-0.10	0.62	0.49	0.42	0.40	0.58
Non-Urban	-0.54	0.18	-0.71	0.29	0.70	0.69	1.18	0.69
Non-owner								
Urbanized Areas 250,000+	1.16	2.56	2.33	3.18	1.39	1.48	1.61	1.62
Other Urban	3.41	3.20	1.23	4.49	1.51	1.74	1.09	1.34
Non-Urban	6.52	6.23	2.85	6.08	4.20	1.71	1.51	1.81
Black								
Owner								
Urbanized Areas 250,000+	1.63	2.16	0.81	6.10	1.91	0.90	0.87	1.91
Other Urban		1.34				0.98		
Non-Urban		3.52				1.90		
Non-owner								
Urbanized Areas 250,000+	8.37	6.27	5.99	9.96	1.61	1.90	1.68	2.72
Other Urban		4.15				1.18		
Non-Urban		4.62				5.33		
Non-Black Hispanic								
Owner								
Urbanized Areas 250,000+	0.67	2.53	-4.33	2.89	4.45	0.90	2.58	0.87
Other Urban		0.94				1.64		
Non-Urban		2.73				2.69		
Non-owner								
Urbanized Areas 250,000+	6.72	9.34	6.64	5.91	3.51	2.59	3.26	1.84
Other Urban		6.60						
Non-Urban		15.80				5.01		
Asian and Pacific Islander								
Owner		-1.45						
Non-owner		6.96				2.52		
American Indians on Reservations		12.22				4.73		

Negative numbers in table signify as estimated overcount.

AREA	LEVEL OF ERROR OVER 10 YEARS[13]
States	1.5 - 2.5%
Places over 50,000	4.0%
Places 5,000 to 50,000	7.0 - 8.0%
Places under 5,000	16.0 - 20.0%

ATTACHMENT 9: USES OF INTERCENSAL ESTIMATES AND ISSUES CONSIDERED BY C.A.P.E.

Uses of Intercensal Estimates:

1. Survey controls
2. Denominators for per capita Federal statistics
3. Funding programs

   A. State populations either for direct funding or as the first tier in a funding program

   B. Substate areas of 100,000 population or larger

   C. Substate areas below 100,000 population

Other Concerns:

1. National population estimates
2. Differential undercount and the perception of fairness
3. Overall accuracy

[13] Level of error as measured in previous decades. These error estimates exclude any estimated undercoverage in the census.

Attachment 11: Place Level Estimates and Estimated Undercount Rates (Places with 100,000 or More Population)

State/	Place/	Place Name	1990 Census	Original PES July 1991	Estimated UCRt	SE(UCRt)	357 PES July 1992	Estimated UCRt	SE(UCRt)
01	0185	Birmingham City	265968	278776	4.594	0.504	273918	2.902	0.750
01	0935	Huntsville City	159789	165498	3.450	0.557	162535	1.689	0.587
01	1165	Mobile City	196278	203932	3.753	0.522	201181	2.437	0.619
01	1180	Montgomery City	187106	194786	3.943	0.516	190738	1.904	0.521
02	0140	Anchorage City	226338	231238	2.119	0.671	232174	2.514	0.518
04	0140	Glendale City	148134	151575	2.270	0.663	150735	1.725	0.371
04	0215	Mesa City	288091	296297	2.770	0.583	292643	1.556	0.638
04	0260	Phoenix City	983403	1013566	2.976	0.569	1003800	2.032	0.515
04	0305	Scottsdale City	130069	132778	2.040	0.589	131178	0.846	0.612
04	0360	Tempe City	141865	147232	3.645	0.588	145453	2.467	0.791
04	0380	Tucson City	405390	419413	3.344	0.577	415971	2.544	0.542
05	1195	Little Rock City	175795	181658	3.228	0.496	179875	2.268	0.610
06	0070	Anaheim City	266406	277711	4.071	0.530	273740	2.679	0.538
06	0180	Bakersfield City	174820	179683	2.706	0.574	179398	2.552	0.511
06	0245	Berkeley City	102724	107538	4.477	0.487	106630	3.664	0.712
06	0525	Chula Vista City	135163	140021	3.470	0.584	138715	2.561	0.475
06	0595	Concord City	111348	113121	1.567	0.622	113137	1.582	0.580
06	0880	El Monte City	106209	112288	5.414	0.745	110792	4.137	0.614
06	0935	Esdondido City	108653	112428	3.374	0.533	111040	2.166	0.549
06	1080	Fremont City	173339	177040	2.091	0.584	176094	1.565	0.522
06	1090	Fresno City	354202	369030	4.018	0.497	366527	3.363	0.555
06	1095	Fullerton City	114144	166779	2.256	0.583	116725	2.211	0.514
06	1110	Garden Grove City	143050	146505	2.358	0.572	146412	2.296	0.515
06	1130	Glendale City	180038	183360	1.812	0.584	184515	2.426	0.579
06	1225	Hayward City	111498	115752	3.675	0.566	114720	2.809	0.503
06	1300	Huntington Beach City	181519	183976	1.336	0.632	184639	1.690	0.635
06	1340	Ingelwood City	109602	123350	11.146	0.953	116991	6.316	1.290
06	1347	Irvine City	110330	111773	1.291	0.631	112191	1.659	0.665
06	1610	Long Beach City	429433	450964	4.774	0.466	445925	3.698	0.594
06	1630	Los Angeles City	3485398	3671205	5.061	0.514	3624206	3.830	0.651
06	1790	Modesto City	164730	168273	2.106	0.601	168849	2.440	0.500
06	1849	Moreno Valley City	118779	126583	6.165	0.563	121925	2.580	0.457
06	1970	Oakland City	372242	392769	5.226	0.540	391553	4.932	0.919

06	1990	Oceanside City	128398	132708	3.248	0.586	131771	2.515	0.510
06	2005	Ontario City	133179	141469	5.860	0.577	137458	3.113	0.551
06	2015	Orange City	110658	113020	2.090	0.590	112738	1.845	0.495
06	2050	Oxnard City	142216	148120	3.986	0.581	147164	3.362	0.643
06	2125	Pasadena City	131591	137947	4.608	0.460	136431	3.548	0.582
06	2230	Pomona City	131723	138469	4.872	0.536	137116	3.933	0.693
06	2278	Rancho Cucamonge City	101409	106655	4.919	0.548	103309	1.839	0.485
06	2370	Riverside City	226505	233085	2.823	0.562	232608	2.624	0.492
06	2420	Sacramento City	369365	384466	3.928	0.477	380736	2.987	0.538
06	2435	Salinas City	108777	113243	3.944	0.595	112703	3.484	0.993
06	2450	San Bernardino City	164164	170413	3.667	0.524	170249	3.574	0.577
06	2475	San Diego City	1110549	1156224	3.950	0.476	1143032	2.842	0.527
06	2485	San Fransisco City	723959	756182	4.261	0.504	745573	2.899	0.626
06	2510	San Jose City	782248	814783	3.993	0.520	801296	2.377	0.474
06	2570	Santa Ana City	293742	309907	5.216	0.648	305815	3.948	0.871
06	2583	Santa Clarita City	110642	112528	1.676	0.647	111997	1.210	0.558
06	2615	Santa Rosa City	113313	115042	1.503	0.668	115898	2.231	0.533
06	2702	Simi Valley City	100217	104425	4.030	0.566	102006	1.754	0.449
06	2805	Stockton City	210943	218902	3.636	0.540	218358	3.396	0.600
06	2835	Sunnyvale City	117229	119490	1.892	0.578	119999	2.308	0.610
06	2897	Thousand Oaks City	104352	108398	3.733	0.565	105407	1.001	0.553
06	2910	Torrance City	133107	134632	1.133	0.601	135125	1.494	0.564
06	3000	Vallejo City	109199	113359	3.670	0.550	112178	2.656	0.544
08	0055	Aurora City	222103	227295	2.284	0.673	227110	2:205	0.583
08	0240	Colorado Springs City	281140	289844	3.003	0.572	287033	2.053	0.635
08	0320	Denver City	467610	482714	3.129	0.579	480862	2.756	0.498
08	0760	Lakewood City	126481	128314	1.429	0.680	128094	1.259	0.649
09	001010	Bridgeport Town	141686	143879	1.524	0.857	145631	2.709	1.029
09	001090	Stamford Town	108056	108286	0.212	0.770	109430	1.256	0.461
09	003070	Hartford Town	139739	143285	2.475	0.957	146308	4.490	1.231
09	009075	New Haven Town	130474	132416	1.467	0.844	135057	3.393	0.842
09	009120	Waterbury Town	108961	109092	0.120	0.759	110722	1.591	0.534
11	0005	Washington City	606900	638747	4.986	0.517	628309	3.407	0.901
12	0645	Fort Lauderdale City	149377	153932	2.959	0.490	152687	2.168	0.660
12	0860	Hialeah City	188004	196416	4.283	0.935	197448	4.783	1.621
12	0915	Hollywood City	121697	125104	2.723	0.509	123463	1.431	0.569
12	1003	Jacksonville City (remainder)	635230	658739	3.569	0.462	649437	2.188	0.548

12	1370	Miami City	358548	376424	4.749	0.703	377379	4.990	1.527
12	1600	Orlando City	164693	170303	3.294	0.462	169260	2.698	0.700
12	1900	St. Petersburg City	238629	245561	2.823	0.472	242149	1.454	0.555
12	2070	Tallahassee City	124773	129647	3.759	0.526	127834	2.395	0.816
12	2075	Tampa City	280015	291356	3.893	0.449	287445	2.585	0.627
13	0150	Atlanta City	394017	415204	5.103	0.540	407923	3.409	0.912
13	0660	Columbus City (remainder)	178681	184860	3.343	0.505	182489	2.087	0.554
13	1725	Macon City	106612	110227	3.280	0.542	109027	2.215	0.586
13	2540	Savannah City	137560	142220	3.277	0.531	140538	2.119	0.560
15	0110	Honolulu CDP	365272	382505	4.505	0.803	372146	1.847	0.989
16	0090	Boise City	125738	127612	1.469	0.702	128336	2.024	0.542
17	1051	Chicago City	2783726	2857364	2.577	0.582	2852041	2.395	0.769
17	4590	Peoria City	113504	116740	2.772	0.681	114753	1.089	0.416
17	4965	Rockford City	139426	143232	2.657	0.681	140598	0.834	0.422
17	5480	Springfield City	105227	107883	2.462	0.700	105921	0.655	0.456
18	0775	Evansville City	126272	129192	2.260	0.712	126950	0.534	0.475
18	0825	Fort Wayne City	173072	177949	2.741	0.690	174511	0.824	0.429
18	0905	Gary City	116646	122166	4.518	0.866	119611	2.479	0.719
18	1145	Indianapolis	731327	737483	0.835	0.612	741712	1.400	0.523
18	2375	South Bend City	105511	108564	2.812	0.681	106417	0.851	0.377
19	0670	Cedar Rapids City	108751	110887	1.926	0.648	109199	0.410	0.430
19	1130	DES Moines City	193187	197761	2.313	0.631	194978	0.919	0.506
20	1430	Kansas City	149767	153306	2.309	0.483	151947	1.435	0.494
20	2194	Overland Park City	111790	112871	0.958	0.491	112485	0.618	0.480
20	2795	Topeka City	119883	123028	2.556	0.602	120748	0.716	0.434
20	3040	Wichita City	304011	308747	1.534	0.480	307807	1.233	0.518
21	1160	Lexington Fayette	225366	233157	3.342	0.602	229930	1.985	0.705
21	1230	Louisville City	269063	279912	3.876	0.499	274816	2.094	0.616
22	0095	Baton Rouge City	219531	227504	3.505	0.479	226061	2.889	0.704
22	0956	New Orleans City	496938	514558	3.424	0.486	513936	3.307	0.876
22	1240	Shreveport City	198525	205361	3.329	0.482	203753	2.566	0.633
24	0025	Baltimore City	736014	772082	4.672	0.511	759127	3.045	0.808
25	013090	Springfield City	156983	158023	0.658	0.785	159597	1.638	0.850
25	017130	Lowell City	103439	103118	0.311	0.770	105772	2.206	0.667
25	025005	Boston City	574283	579743	0.942	0.806	590703	2.780	0.784
25	027300	Worcester City	169759	169075	0.405	0.753	171148	0.812	0.816
26	0080	Ann Arbor City	109592	112804	2.847	0.727	111442	1.660	0.522
26	0680	Detroit City	1027974	1064760	3.455	0.622	1056180	2.671	0.727
26	0920	Flint City	140761	146209	3.726	0.703	143923	2.197	0.584

26	1085	Grand Rapids City	189126	194874	2.950	0.666	191834	1.412	0.504
26	1485	Lansing City	127321	131473	3.158	0.684	129424	1.625	0.553
26	1565	Livonia City	100850	101462	0.603	0.527	100700	0.149	0.364
26	2583	Sterling Heights City	117810	118625	0.687	0.514	117955	0.123	0.402
26	2790	Warren City	144864	145814	0.652	0.535	145018	0.106	0.394
27	2585	Minneapolis City	368383	374965	1.755	0.469	374537	1.643	0.605
27	3425	St. Paul City	272235	275845	1.309	0.485	275962	1.351	0.560
28	0615	Jackson City	196637	205662	4.388	0.515	202591	2.939	0.719
29	2125	Independence City	112301	113335	0.912	0.487	112970	0.592	0.493
29	2220	Kansas City	435146	444859	2.183	0.472	441627	1.468	0.516
29	3875	St. Louis City	396685	408263	2.836	0.518	405175	2.096	0.682
29	4075	Springfield City	140494	143438	2.053	0.650	141440	0.669	0.501
31	1425	Lincoln City	191972	196234	2.172	0.660	193365	0.720	0.455
31	1825	Omaha City	335795	340507	1.384	0.476	339436	1.073	0.498
32	0065	Las Vegas City	258295	266308	3.009	0.562	264680	2.412	0.535
32	0090	Reno City	133850	136305	1.801	0.650	137829	2.887	0.670
34	1715	Elizabeth City	110002	111988	1.773	0.740	113626	3.189	1.244
34	2290	Jersey City	228537	236712	3.454	0.681	236914	3.536	0.942
34	2895	Newark City	275221	285923	3.743	0.775	289965	5.085	1.113
34	3115	Paterson City	140891	146967	4.134	0.752	146865	4.068	1.332
35	0015	Albuquerque City	384736	397206	3.139	0.583	393462	2.218	0.480
36	0030	Albany City	101082	103456	2.295	0.692	103108	1.965	0.802
36	0450	Buffalo City	328123	333145	1.508	0.592	334286	1.844	0.726
36	2505	New York City	7322564	7552196	3.041	0.588	7567146	3.232	0.921
36	3100	Rochester City	231636	239832	3.417	0.720	237133	2.318	0.746
36	3565	Syracuse City	163860	167479	2.161	0.683	166653	1.676	0.769
36	4075	Yonkers City	188082	192435	2.262	0.664	190656	1.350	0.852
37	0480	Charlotte City	395934	412466	4.008	0.467	405932	2.463	0.635
37	0750	Durham City	136611	141713	3.600	0.536	139962	2.394	0.712
37	1065	Greensboro City	183521	189851	3.334	0.518	187128	1.928	0.646
37	2020	Raleigh City	207951	215573	3.536	0.520	213485	2.592	0.728
37	2785	Winston-Salem City	143485	148215	3.191	0.513	146388	1.983	0.619
39	0035	Akron City	223019	229527	2.835	0.683	226256	1.431	0.520
39	0865	Cincinnati City	364040	369165	1.388	0.631	372392	2.243	0.719
39	0900	Cleveland City	505616	512581	1.359	0.637	516598	2.126	0.650
39	0960	Columbus City	632910	639303	1.000	0.605	645256	1.913	0.630
39	1110	Dayton City	182044	188260	3.302	0.670	185861	2.054	0.624
39	4265	Toledo City	332943	335164	0.663	0.600	337317	1.297	0.497
40	1815	Oklahoma City	444719	454958	2.251	0.516	454630	2.180	0.548

40	2465	Tulsa City	367302	375358	2.146	0.539	374856	2.015	0.597
41	0360	Eugene City	112669	114413	1.524	0.702	115726	2.641	0.685
41	0905	Portland City	437319	450413	2.907	0.538	445566	1.851	0.659
41	1005	Salem City	107786	109189	1.285	0.652	110240	2.227	0.546
42	0165	Allentown City	105090	105902	0.767	0.627	105216	0.120	0.831
42	3685	Erie City	108718	110075	1.233	0.662	109866	1.045	0.534
42	7180	Philadelphia City	1585577	1606249	1.287	0.609	1608942	1.452	0.742
42	7234	Pittsburgh City	369879	374002	1.102	0.583	373752	1.036	0.728
44	007065	Providence City	160728	161519	0.490	0.777	164304	2.176	0.829
46	1225	Sioux Falls City	100814	102712	1.848	0.658	101208	0.389	0.496
47	0245	Chattanooga City	152466	157807	3.385	0.528	155875	2.187	0.637
47	0760	Knoxville City	165121	170454	3.129	0.587	168582	2.053	0.698
47	0940	Memphis City	610337	640010	4.636	0.498	628329	2.864	0.709
47	1016	Nashville-Davidson (remainder)	488374	508302	4.109	0.519	499383	2.205	0.625
48	0015	Abilene City	106654	109869	2.926	0.515	108885	2.049	0.646
48	0100	Amarillo City	157615	162215	2.836	0.532	160530	1.816	0.577
48	0175	Arlington City	261721	272160	3.836	0.510	269098	2.742	0.608
48	0210	Austin City	465622	480242	3.044	0.501	483156	3.629	0.752
48	0320	Beaumont City	114323	118161	3.248	0.474	116654	1.998	0.500
48	0980	Corpus Christi City	257453	264658	2.722	0.551	267127	3.622	0.798
48	1085	Dallas City	1006877	1057658	4.801	0.508	1043947	3.551	0.727
48	1340	El Paso City	515342	531606	3.059	0.637	538250	4.256	0.964
48	1500	Fort Worth City	447619	467853	4.325	0.490	461686	3.047	0.606
48	1580	Garland City	180650	185940	2.845	0.494	185336	2.528	0.539
48	1975	Houston City	1630553	1715633	4.959	0.542	1697301	3.933	0.777
48	2060	Irving City	155037	162091	4.352	0.530	160622	3.477	0.762
48	2400	Laredo City	122899	127296	3.454	0.793	126611	2.932	1.262
48	2565	Lubbock City	186206	192375	3.207	0.512	190661	2.336	0.688
48	2795	Mesquite City	101484	104448	2.838	0.503	103803	2.234	0.541
48	3200	Pasadena City	119363	123270	3.170	0.588	123539	3.380	0.721
48	3310	Plano City	128713	132377	2.768	0.519	131188	1.887	0.540
48	3745	San Antonio City	935933	964071	2.919	0.561	974099	3.918	0.857
48	4415	Waco City	103590	107015	3.201	0.476	106382	2.624	0.728
49	0870	Salt Lake City	159936	162897	1.818	0.664	163014	1.888	0.721
51	0025	Alexandria City	111183	112748	1.388	0.541	114451	2.856	0.771
51	0242	Chesapeake City	151976	153512	1.001	0.556	155185	2.068	0.509
51	0590	Hampton City	133793	139284	3.942	0.459	137415	2.636	0.617
51	0860	Newport News City	170045	178053	4.498	0.468	175121	2.899	0.689
51	0875	Norfolk City	261229	273457	4.472	0.444	269011	2.893	0.733

51	0990	Portsmouth City	103907	108477	4.213	0.474	106837	2.742	0.695
51	1035	Richmond City	203056	209959	3.288	0.549	208987	2.838	0.817
51	1280	Virginia Beach City	393069	408213	3.710	0.487	402092	2.244	0.558
53	1140	Seattle City	516259	534576	3.427	0.506	528151	2.252	0.670
53	1220	Spokane City	177196	179308	1.178	0.711	179391	1.223	0.739
53	1280	Tacoma City	176664	180714	2.241	0.625	180831	2.304	0.622
55	1475	Madison City	191262	196296	2.565	0.734	193499	1.156	0.504
55	1645	Milwaukee City	628088	635933	1.234	0.601	642860	2.298	0.681

UCRT Undercount Rate as estimatred etc.

Attachment 12: County Level Estimates and Estimated Undercount Rates (counties with 100,000 or more population)

State	County	County Name	1990 Census	Original PES July 1991 Estimate UC Rt SE(UCRt)			357 PES July 1992 Estimate UC Rt SE(UCRt)		
01	015	Calhoun County	116034	119037	2.523	0.466	117856	1.546	0.424
01	073	Jefferson County	651525	673700	3.292	0.423	665329	2.075	0.517
01	089	Madison County	238912	246704	3.158	0.456	242937	1.657	0.478
01	097	Mobile County	378643	390685	3.082	0.417	387137	2.194	0.479
01	101	Montgomery County	209085	217215	3.743	0.481	213105	1.886	0.480
01	103	Morgan County	100043	102781	2.664	0.459	101438	1.375	0.407
01	125	Tuscaloosa County	150522	155432	3.159	0.424	153449	1.908	0.508
02	020	Anchorage Borough	226338	231238	2.119	0.671	232174	2.514	0.518
04	013	Maricopa County	2122101	2180538	2.680	0.496	2160697	1.786	0.512
04	019	Pima County	666880	686848	2.907	0.486	681920	2.206	0.464
04	021	Pinal County	116379	121955	4.572	0.577	120033	3.045	0.584
04	025	Yavapai County	107714	110720	2.715	0.575	109685	1.797	0.442
04	027	Yuma County	106895	111958	4.522	0.570	109960	2.787	0.572
05	119	Pulaski County	349660	360243	2.938	0.432	357441	2.177	0.517
05	143	Washington County	113409	116428	2.593	0.474	115578	1.877	0.615
06	001	Alameda County	1279182	1323971	3.383	0.455	1317233	2.889	0.5xx
06	007	Butte County	182120	187906	3.079	0.548	186831	2.522	0.554
06	013	Contra Costa County	803732	825024	2.581	0.603	817943	1.737	0.400
06	017	El Dorado County	125995	126797	0.633	0.696	128413	1.883	0.451
06	019	Fresno County	667490	692890	3.666	0.457	691987	3.540	0.501
06	023	Humbolt County	119118	122156	2.487	0.582	122410	2.689	0.488
06	025	Imperial County	109303	116024	5.793	0.705	113220	3.460	0.866
06	029	Kern County	543477	566235	4.019	0.473	558755	2.734	0.375
06	031	Kings County	101469	105597	3.909	0.504	105099	3.454	0.581
06	037	Los Angeles County	8863164	9291955	4.615	0.448	9168889	3.334	0.548
06	041	Marin County	230096	232036	0.836	0.651	232947	1.224	0.523
06	047	Merced County	178403	186707	4.448	0.470	185406	3.777	0.628
06	053	Monterey County	355660	370124	3.908	0.441	367580	3.243	0.644
06	055	Napa County	110765	113411	2.333	0.503	113298	2.236	0.447
06	059	Orange County	2410556	2469336	2.380	0.519	2461373	2.065	0.493
06	061	Placer County	172796	174772	1.131	0.575	175303	1.430	0.374
06	065	Riverside County	1170413	1220764	4.125	0.487	1198964	2.381	0.343
06	067	Sacramento County	1041219	1069918	2.682	0.491	1065198	2.251	0.524
06	071	San Bernardino County	1418380	1490697	4.851	0.501	1455550	2.554	0.355
06	073	San Diego County	2498016	2576888	3.061	0.442	2560392	2.436	0.486
06	075	San Francisco County	723959	756182	4.261	0.504	745573	2.899	0.626

06	077	San Joaquin County	480628	498718	3.627	0.453	495154	2.934	0.381
06	079	San Luis Obispo County	217162	222991	2.614	0.513	222841	2.549	0.500
06	081	San Mateo County	649623	664465	2.234	0.571	561709	1.826	0.457
06	083	Santa Barbara County	369608	383034	3.505	0.473	381039	3.000	0.645
06	085	Santa Clara County	1497577	1544157	3.017	0.453	1531196	2.196	0.475
06	087	Santa Cruz County	229734	238267	3.581	0.503	236007	2.658	0.531
06	089	Shasta County	147036	150573	2.349	0.528	150145	2.070	0.447
06	095	Solano County	340421	353913	3.812	0.450	348512	2.322	0.324
06	097	Sonoma County	388222	399078	2.720	0.504	397377	2.304	0.422
06	099	Stanislaus County	370522	382342	3.092	0.487	380699	2.673	0.475
06	107	Tulane County	311921	324294	3.815	0.482	323520	3.585	0.681
06	111	Ventura County	669016	694637	3.688	0.468	683672	2.144	0.357
06	113	Yolo County	141092	145883	3.284	0.456	145975	3.345	0.565
08	001	Adams County	265038	271716	2.458	0.654	269856	1.786	0.496
08	005	Arapahoe County	391511	398166	1.671	0.693	397542	1.517	0.588
08	013	Boulder County	225339	229447	1.790	0.591	230754	2.347	0.533
08	031	Denver County	467610	482714	3.129	0.579	480862	2.756	0.498
08	041	El Paso County	397014	407843	2.655	0.493	405212	2.023	0.558
08	059	Jefferson County	438430	444327	1.327	0.706	442890	1.007	0.577
08	069	Larimer County	186136	189346	1.695	0.596	190569	2.326	0.527
08	101	Pueblo County	123051	125654	2.072	0.550	125754	2.149	0.540
08	123	Weld County	131821	134887	2.273	0.534	135793	2.925	0.518
09	001	Fairfield County	827645	831105	0.416	0.593	832682	0.605	0.384
09	003	Hartford County	851783	857182	0.630	0.589	857897	0.713	0.483
09	005	Litchfield County	174092	175581	0.848	0.538	175080	0.565	0.523
09	007	Middlesex County	143196	143812	0.428	0.529	143825	0.437	0.526
09	009	New Haven County	804219	807947	0.461	0.583	807987	0.466	0.514
09	011	New London County	254957	255796	0.328	0.554	257535	1.001	0.470
09	013	Tolland County	128699	129683	0.759	0.599	129510	0.626	0.561
09	015	Windham County	102525	104554	1.941	0.823	103793	1.222	0.681
10	001	Kent County	110993	114068	2.696	0.443	112995	1.772	0.394
10	003	New Castle County	441946	456338	3.154	0.510	450294	1.854	0.516
10	005	Sussex County	113229	116255	2.603	0.501	115083	1.611	0.452
11	001	District of Columbia	606900	638747	4.986	0.517	628309	3.407	0.901
12	001	Alachua County	181596	188223	3.521	0.429	186051	2.394	0.635
12	005	Bay County	126994	130912	2.993	0.477	129096	1.629	0.536
12	009	Brevard County	398978	410499	2.807	0.446	404953	1.476	0.445
12	011	Broward County	1255488	1291812	2.812	0.453	1277394	1.715	0.529
12	015	Charlotte County	110975	112871	1.680	0.526	111898	0.825	0.353

12	019	Clay County	105986	106804	0.766	0.595	107762	1.648	0.376
12	021	Collier County	152099	156294	2.684	0.526	154958	1.845	0.464
12	025	Dade County	1937094	1997643	3.031	0.591	2011300	3.690	0.945
12	031	Duval County	672971	697735	3.549	0.463	687821	2.159	0.549
12	033	Escambia County	262798	271007	3.029	0.466	268329	2.061	0.495
12	053	Hernando County	101115	100975	-0.139	0.612	102051	0.918	0.319
12	057	Hillsborough County	834054	851877	2.092	0.448	853411	2.268	0.478
12	069	Lake County	152104	155095	1.929	0.481	154003	1.233	0.341
12	071	Lee County	335113	343538	2.452	0.465	339589	1.318	0.466
12	073	Leon County	192493	199708	3.613	0.437	196621	2.100	0.615
12	081	Manatee County	211707	216819	2.358	0.508	214609	1.352	0.513
12	083	Marion County	194833	199845	2.508	0.487	197743	1.472	0.354
12	085	Martin County	100900	103232	2.259	0.592	102120	1.195	0.406
12	091	Okaloosa County	143776	148410	3.122	0.505	146346	1.756	0.593
12	095	Orange County	677491	700574	3.295	0.458	693622	2.326	0.530
12	097	Osceola County	107728	111188	3.112	0.564	109720	1.816	0.479
12	099	Palm Beach County	863518	886676	2.612	0.484	876764	1.511	0.493
12	101	Pasco County	281131	281049	-0.029	0.614	283694	0.904	0.395
12	103	Pinellas County	851659	861306	1.120	0.448	860438	1.020	0.555
12	105	Polk County	405382	416923	2.768	0.470	411918	1.587	0.405
12	111	St. Lucie County	150171	154362	2.715	0.479	152554	1.562	0.474
12	115	Sarasota County	277776	283554	2.038	0.550	279921	0.766	0.505
12	117	Seminole County	287529	297007	3.191	0.569	292736	1.779	0.505
12	127	Volusia County	370712	380601	2.598	0.512	375737	1.338	0.463
13	021	Bibb County	149967	154963	3.224	0.453	157035	2.005	0.475
13	051	Chatham County	216935	224122	3.207	0.435	221102	1.885	0.506
13	063	Clayton County	182052	184137	1.132	0.562	186841	2.563	0.581
13	067	Cobb County	447745	453535	1.277	0.544	456480	1.914	0.547
13	089	DeKalb County	545837	553706	1.421	0.533	561155	2.730	0.608
13	121	Fulton County	648951	671488	3.356	0.442	668695	2.953	0.738
13	135	Gwinnett County	352910	356619	1.040	0.611	359473	1.826	0.488
13	215	Muscogee County	179278	185474	3.341	0.505	183097	2.086	0.554
13	245	Richmond County	189719	195914	3.162	0.443	194873	2.645	0.584
15	001	Hawaii County	120317	121720	1.153	0.717	122654	1.905	0.750
15	003	Honolulu County	836231	861245	2.904	0.570	852074	1.859	0.837
15	009	Maui County	100374	101591	1.198	0.714	102187	1.774	0.741
16	001	Ada County	205775	208426	1.272	0.594	209575	1.813	0.463
17	019	Champaign County	173025	177031	2.263	0.553	175375	1.340	0.414
17	031	Cook County	5105067	5212195	2.055	0.423	5186429	1.569	0.574

17	043	DuPage County	781666	789453	0.986	0.499	784956	0.419	0.399
17	089	Kane County	317471	324570	2.187	0.524	320253	0.869	0.413
17	097	Lake County	516418	524672	1.573	0.558	519660	0.624	0.330
17	099	LaSalle County	106913	106411	0.472	0.538	107150	0.222	0.416
17	111	McHenry County	183241	184777	0.831	0.510	183780	0.293	0.397
17	113	McLean County	129180	131827	2.008	0.582	130128	0.729	0.408
17	115	Macon County	117206	119550	1.961	0.570	117856	0.551	0.357
17	119	Madison County	249238	251156	0.764	0.432	250446	0.483	0.305
17	143	Peoria County	182827	186534	1.987	0.534	184180	0.735	0.372
17	161	Rock Island County	148723	151424	1.784	0.534	149787	0.711	0.451
17	163	St. Clair County	262852	266701	1.443	0.423	266421	1.340	0.409
17	167	Sangamon County	178386	181578	1.758	0.542	179149	0.426	0.399
17	179	Tazwell County	123692	124872	0.945	0.561	123942	0.202	0.407
17	197	Will County	357313	363530	1.710	0.554	359200	0.525	0.284
17	201	Winnebago County	252913	257702	1.858	0.528	254302	0.546	0.378
18	003	Allen County	300836	306760	1.931	0.534	302274	0.476	0.392
18	035	Delaware County	119659	121730	1.701	0.537	120341	0.566	0.402
18	039	Elkhart County	156198	158664	1.554	0.530	156797	0.382	0.443
18	057	Hamilton County	108936	109674	0.673	0.513	109211	0.252	0.385
18	089	Lake County	475594	487249	2.392	0.552	480322	0.984	0.427
18	091	LaPorte County	107066	107036	-0.028	0.462	107368	0.281	0.480
18	095	Madison County	130669	132535	1.408	0.514	131090	0.321	0.403
18	097	Marion County	797159	803890	0.837	0.577	808143	1.359	0.523
18	127	Porter County	128932	130035	0.848	0.659	129287	0.274	0.397
18	105	Monroe County	108928	111084	1.896	0.552	110094	-1.013	0.498
18	141	St. Joseph County	247052	251786	1.880	0.535	248403	0.544	0.355
18	157	Tippecanoe County	130598	133031	1.829	0.550	132098	1.135	0.459
18	163	Vanderburgh County	165058	168249	1.897	0.596	165711	0.394	0.418
18	167	Vigo County	106107	107712	1.490	0.517	106607	0,469	0.398
19	013	Black Hawk County	123798	126453	2.100	0.553	124529	0.587	0.373
19	113	Linn County	168767	171900	1.823	0.541	169329	0.332	0.387
19	153	Polk County	327140	334027	2.062	0.537	329530	0.725	0.432
19	163	Scott County	150979	154206	2.093	0.533	152246	0.832	0.431
20	091	Johnson County	355054	358386	0.930	0.435	357029	0.553	0.418
20	173	Sedgwick County	403662	409349	1.389	0.407	407780	1.010	0.440
20	177	Shawnee County	160976	164773	2.304	0.525	161845	0.537	0.394
20	209	Wyandotte County	161993	165674	2.222	0.456	164206	1.348	0.460
21	067	Fayette County	225366	233157	3.342	0.602	229930	1.985	0.705
21	111	Jefferson County	664937	685007	2.930	0.439	676776	1.749	0.537

21	117	Kenton County	142031	145523	2.400	0.593	144235	1.528	0.552
22	017	Caddo Parish	248253	256120	3.072	0.428	254356	2.400	0.529
22	019	Calcasieu Parish	168134	172829	2.717	0.405	170974	1.661	0.420
22	033	East Baton Rouge Parish	380105	392277	3.103	0.395	390145	2.574	0.569
22	051	Jefferson Parish	448306	458990	2.326	0.470	457937	2.103	0.525
22	055	Lafayette Parish	164762	169813	2.974	0.409	168125	2.000	0.497
22	071	Orleans Parish	496938	514558	3.424	0.486	513936	3.307	0.876
22	073	Quachita Parish	142191	146297	2.807	0.400	144953	1.905	0.438
22	079	Rapides Parish	131556	135085	2.612	0.389	133995	1.820	0.399
22	103	St. Tammany Parish	144508	147804	2.230	0.451	146874	1.611	0.365
23	001	Androscoggin County	105259	104912	-0.331	0.585	106120	0.812	0.511
23	005	Cumberland County	243135	243615	0.197	0.539	245246	0.861	0.524
23	011	Kennebec County	115904	117501	1.359	0.693	116582	0.581	0.505
23	019	Penobscot County	146601	147574	0.659	0.563	147738	0.770	0.532
23	031	York County	164587	166105	0.914	0.552	165635	0.633	0.520
24	003	Anne Arundel County	427239	431624	1.016	0.537	434447	1.659	0.406
24	005	Baltimore County	692134	696225	0.588	0.567	702812	1.519	0.507
24	013	Carroll County	123372	124098	0.585	0.606	124911	1.232	0.459
24	017	Charles County	101154	102192	1.016	0.571	102794	1.595	0.403
24	021	Frederick County	150208	152604	1.570	0.494	152690	1.626	0.431
24	025	Hartford County	182132	183499	0.745	0.583	185018	1.560	0.359
24	027	Howard County	187328	189033	0.902	0.582	190409	1.618	0.466
24	031	Montgomery County	757027	764514	0.979	0.563	771160	1.833	0.482
24	033	Prince Georges County	729268	740060	1.458	0.579	751587	2.970	0.627
24	043	Washington County	121393	124802	2.732	0.464	123237	1.496	0.460
24	510	Baltimore City	736014	772082	4.672	0.511	759127	3.045	0.808
25	001	Barnstable County	186605	189889	1.729	0.855	187904	0.691	0.530
25	003	Berkshire County	139352	139722	0.265	0.520	140508	0.823	0.505
25	005	Bristol County	506325	505255	-0.212	0.554	509637	0.650	0.452
25	009	Essex County	670080	670474	0.059	0.579	671451	0.204	0.466
25	013	Hampden County	456310	457899	0.347	0.585	458054	0.381	0.706
25	015	Hampshire County	146568	147943	0.929	0.563	147848	0.866	0.555
25	017	Middlesex County	1398468	1402907	0.316	0.600	1399207	0.053	0.615
25	021	Norfolk County	616087	618087	0.324	0.653	611139	-0.810	0.744
25	023	Plymouth County	435276	436386	0.254	0.580	436400	0.258	0.406
25	025	Suffolk County	663906	670095	0.924	0.744	680818	2.484	0.777
25	027	Worcester County	709705	711256	0.218	0.537	713339	0.509	0.456
26	017	Bay County	111723	113132	1.245	0.537	111895	0.153	0.450
26	021	Berrien County	161378	163661	1.395	0.598	162674	0.796	0.454

26	025	Calhoun County	135982	138148	1.568	0.517	136672	0.505	0.398
26	049	Genesse County	430459	438800	1.901	0.538	434600	0.953	0.414
26	065	Ingham County	281912	288505	2.285	0.534	286089	1.460	0.534
26	075	Jackson County	149756	151533	1.173	0.526	150189	0.288	0.511
26	077	Kalamazoo County	223411	227212	1.673	0.520	224957	0.687	0.406
26	081	Kent County	500631	509273	1.697	0.526	504353	0.738	0.407
26	093	Livingston County	115645	116408	0.656	0.511	115499	-0.126	0.949
26	099	Macomb County	717400	722597	0.719	0.522	718766	0.190	0.387
26	115	Monroe County	133600	134642	0.774	0.511	133783	0.137	0.577
26	121	Muskegon County	158983	161494	1.555	0.535	159784	0.501	0.394
26	125	Oakland County	1083592	1094932	1.036	0.481	1088374	0.439	0.383
26	139	Ottawa County	187768	189955	1.151	0.605	188460	0.367	0.443
26	145	Saginaw County	211946	216155	1.947	0.537	213567	0.759	0.401
26	147	St. Clair County	145607	147341	1.177	0.440	145854	0.169	0.512
26	161	Washtenaw County	282937	288679	1.989	0.516	286038	1.084	0.427
26	163	Wayne County	2111687	2160354	2.253	0.426	2144482	1.529	0.478
27	003	Anoka County	243641	245862	0.903	0.517	244251	0.250	0.375
27	037	Dakota County	275227	278038	1.011	0.512	276471	0.450	0.389
27	053	Hennepin County	1032431	1044852	1.189	0.381	1041265	0.848	0.467
27	109	Olmsted County	106470	108411	1.790	0.553	106753	0.265	0.411
27	123	Ramsey County	485765	491319	1.130	0.382	490387	0.943	0.479
27	137	St. Louis County	198213	201605	1.683	0.576	198462	0.126	0.430
27	145	Stearns County	118791	121193	1.982	0.639	119274	0.405	0.560
27	163	Washington County	145896	147156	0.856	0.506	146053	0.108	0.344
28	047	Harrison County	165365	170273	2.882	0.422	168426	1.818	0.509
28	049	Hinds County	254441	264818	3.919	0.446	261731	2.785	0.609
28	059	Jackson County	115243	118271	2.560	0.460	117089	1.576	0.407
29	019	Boone County	112379	115311	2.543	0.550	113620	1.092	0.444
29	047	Clay County	153411	154746	0.863	0.396	154298	0.575	0.414
29	077	Greene County	207949	211970	1.897	0.545	208941	0.475	0.429
29	095	Jackson County	633232	645060	1.834	0.378	640624	1.154	0.466
29	099	Jefferson County	171380	172865	0.859	0.510	171632	0.147	0.504
29	183	St. Charles County	212907	215015	0.980	0.431	213851	0.442	0.380
29	189	St. Louis County	993529	1010023	1.633	0.458	999753	0.623	0.370
29	510	St. Louis City	396685	408263	2.836	0.518	405175	2.096	0.682
30	111	Yellowstone County	113419	114710	1.125	0.605	115539	1.835	0.450
31	055	Douglas County	416444	421918	1.297	0.419	420353	0.930	0.453
31	109	Lancaster County	213641	218226	2.101	0.611	215022	0.642	0.420
31	153	Sarpy County	102583	104050	1.410	0.492	103780	1.154	0.483

32	003	Clark County	741459	759866	2.422	0.518	758692	2.271	0.521
32	031	Washoe County	254667	258898	1.634	0.556	261007	2.429	0.510
33	011	Hillsborough County	336073	335652	-0.125	0.578	338911	0.838	0.500
33	013	Merrimack County	120005	121598	1.310	0.636	120910	0.748	0.539
33	015	Rockingham County	245845	246967	0.454	0.586	247556	0.691	0.546
33	017	Strafford County	104233	104021	-0.204	0.583	105081	0.807	0.557
34	001	Atlantic County	224327	227837	1.541	0.546	226943	1.153	0.374
34	003	Bergen County	825380	829281	0.470	0.580	820928	-0.542	0.786
34	005	Burlington County	395066	401239	1.539	0.665	394939	-0.032	0.568
34	007	Camden County	302824	510058	1.418	0.621	503429	0.120	0.719
34	011	Cumberland County	138053	140210	1.538	0.530	139656	1.148	0.379
34	013	Essex County	778206	802268	2.999	0.560	799678	2.685	0.782
34	015	Glouster County	230082	233020	1.261	0.699	229106	-0.426	0.624
34	017	Hudson County	553099	568477	2.705	0.577	569258	2.839	1.107
34	019	Hunterdon County	107776	107861	0.079	0.603	108451	0.623	0.745
34	021	Mercer County	325824	331440	1.694	0.544	328647	0.859	0.554
34	023	Middlesex County	671780	677682	0.871	0.548	672992	0.180	0.712
34	025	Monmouth County	553124	556412	0.591	0.574	550805	-0.421	0.687
34	027	Morris County	421353	425501	0.975	0.717	419138	-0.529	0.670
34	029	Ocean County	433203	433516	0.072	0.599	429899	-0.769	0.702
34	031	Passaic County	453060	461845	1.902	0.541	459194	1.336	0.858
34	035	Somerset County	240279	241669	0.575	0.578	239512	-0.320	0.617
34	037	Sussex County	130943	132073	0.856	0.729	131218	0.210	0.539
34	039	Union County	493819	503004	1.826	0.588	497433	0.727	0.778
35	001	Bernalillo County	480577	497633	3.427	0.518	491854	2.293	0.457
35	013	Dona Ana County	135510	141574	4.283	0.545	139939	3.165	0.665
36	001	Albany County	292594	295111	0.853	0.530	293849	0.427	0.656
36	005	Bronx County	1203789	1245874	3.378	0.730	1265768	4.897	1.410
36	007	Broome County	212160	212548	0.183	0.541	213689	0.716	0.458
36	013	Chautauqua County	141895	141997	0.072	0.525	143047	0.805	0.539
36	027	Dutchess County	259462	261192	0.662	0.543	261808	0.896	0.459
36	029	Erie County	968532	976594	0.826	0.588	969213	0.070	0.650
36	045	Jefferson County	110943	112132	1.060	0.562	112635	1.503	0.718
36	047	Kings County	2300664	2379894	3.329	0.592	2389150	3.704	0.906
36	055	Monroe County	713968	722929	1.240	0.536	716126	0.301	0.641
36	059	Nassau County	1287348	1296128	0.677	0.571	1277449	-0.775	0.827
36	061	New York County	1487536	1537991	3.281	0.596	1541441	3.497	0.969
36	063	Niagra County	220756	221792	0.467	0.537	220729	-0.012	0.512
36	065	Oneida County	250836	251805	0.385	0.510	252906	0.819	0.447

36	067	Onondaga County	468973	472839	0.818	0.532	469750	0.165	0.638
36	071	Orange County	307647	309752	0.680	0.564	310882	1.040	0.451
36	075	Oswego County	121771	121870	0.081	0.623	122882	0.904	0.685
36	081	Queens County	1951598	2004192	2.624	0.624	1992006	2.029	0.806
36	083	Rensselaer County	154429	154995	0.365	0.535	155072	0.415	0.591
36	085	Richmond County	378977	384245	1.371	0.533	378782	-0.052	0.722
36	087	Rockland County	265475	269627	1.540	0.688	264771	-0.266	0.734
36	089	St. Lawrence County	111974	112733	0.673	0.594	113179	1.064	0.684
36	091	Saratoga County	181276	181488	0.117	0.615	181850	0.316	0.500
36	093	Schenectady County	149285	149852	0.378	0.524	148589	-0.468	0.720
36	103	Suffolk County	1321864	1330743	0.667	0.576	1313346	-0.649	0.727
36	111	Ulster County	165304	167147	1.103	0.612	167385	1.244	0.736
36	119	Westchester County	874866	890648	1.772	0.641	879705	0.550	0.687
37	001	Alamance County	108213	111418	2.877	0.439	109811	1.455	0.408
37	021	Buncombe County	174821	179768	2.752	0.465	177162	1.321	0.413
37	035	Catawba County	118412	112063	2.991	0.498	120094	1.401	0.426
37	051	Cumberland County	274566	284189	3,386	0.419	280604	2.152	0.514
37	057	Davidson County	126677	130509	2.936	0.580	128544	1.453	0.455
37	063	Durham County	181835	188378	3.473	0.462	185785	2.126	0.579
37	067	Forsyth County	265878	274462	3.128	0.430	270363	1.659	0.469
37	071	Gaston County	175093	177824	1.536	0.464	177837	1.543	0.456
37	081	Guilford County	347420	358847	3.184	0.443	353615	1.752	0.501
37	119	Mecklenburg County	511433	528981	3.317	0.424	523306	2.269	0.557
37	129	New Hanover County	120284	124111	3.084	0.438	122381	1.714	0.540
37	133	Onslow County	149838	154392	2.950	0.374	153141	2.157	0.415
37	147	Pitt County	107924	110732	2.536	0.423	110516	2.345	0.557
37	151	Randolph County	106546	109790	2.955	0.595	108009	1.354	0.431
37	155	Robeson County	105179	108097	2.699	0.452	107475	2.136	0.534
37	159	Rowan County	110605	111420	0.732	0.524	112305	1.514	0.375
37	183	Wake County	423380	438428	3.432	0.434	432630	2.138	0.493
37	191	Wayne County	104666	107153	2.321	0.401	106769	1.969	0.390
38	017	Cass County	102874	105012	2.036	0.571	103452	0.559	0.461
39	003	Allen County	109755	111410	1.486	0.510	110262	0.460	0.411
39	017	Butler County	291479	295537	1.373	0.535	292902	0.486	0.359
39	023	Clark County	147548	149800	1.503	0.519	148179	0.426	0.406
39	025	Clermont County	150187	151277	0.721	0.514	150784	0.396	0.522
39	029	Columbia County	108276	107516	-0.679	0.584	108375	0.091	0.584
39	035	Cuyahoga County	1412140	1429431	1.210	0.431	1427932	1.106	0.471
39	045	Fairfield County	103461	103995	0.514	0.427	103594	0.129	0.522

39	049	Franklin County	961437	970249	0.908	0.463	975539	1.446	0.539
39	057	Greene County	136731	138166	1.039	0.632	137700	0.704	0.328
39	061	Hamilton County	866228	876347	1.155	0.424	876795	1.205	0.485
39	085	Lake County	215499	216985	0.685	0.519	216122	0.288	0.378
39	089	Licking County	128300	129042	0.575	0.432	128558	0.201	0.519
39	093	Lorain County	271126	275982	1.760	0.520	272668	0.565	0.364
39	095	Lucas County	462361	465553	0.686	0.477	467096	1.014	0.437
39	099	Mahoning County	264806	268995	1.557	0.528	266443	0.614	0.379
39	103	Medina County	122354	123157	0.652	0.514	122484	0.106	0.462
39	113	Montgomery County	573809	583903	1.729	0.528	580267	1.113	0.461
39	133	Portage County	142585	144241	1.148	0.573	143615	0.717	0.542
39	139	Richland County	126137	127829	1.324	0.520	126535	0.314	0.418
39	151	Stark County	367585	372544	1.331	0.525	368829	0.337	0.384
39	153	Summit County	514990	523958	1.712	0.520	518979	0.769	0.415
39	155	Trumbull County	227813	230339	1.097	0.560	228736	0.403	0.397
39	165	Warren County	113909	114657	0.652	0.498	114158	0.218	0.364
39	169	Wayne County	101461	100828	-0.628	0.605	101745	0.279	0.620
39	173	Wood County	113269	113881	0.537	0.446	113912	0.565	0.418
40	027	Cleveland County	174253	178292	2.265	0.466	177845	2.020	0.539
40	031	Comanche County	111486	114833	2.915	0.418	113756	1.996	0.506
40	109	Oklahoma County	599611	613697	2.295	0.419	612788	2.150	0.547
40	143	Tulsa County	503341	514637	2.195	0.453	512955	1.874	0.534
41	005	Clackamas County	278850	279977	0.403	0.724	281892	1.079	0.452
41	029	Jackson County	146389	150125	2.489	0.537	149287	1.941	0.441
41	039	Lane County	282912	289415	2.247	0.551	289266	2.197	0.493
41	047	Marion County	228483	234494	2.563	0.508	233587	2.185	0.434
41	051	Multnomah County	583887	598049	2.368	0.489	593788	1.668	0.652
41	067	Washington County	311554	314044	0.793	0.688	315806	1.346	0.623
42	003	Allegheny County	1336449	1346520	0.748	0.600	1331707	-0.356	0.758
42	007	Beaver County	186093	186376	0.152	0.593	185256	-0.452	0.637
42	011	Berks County	336523	337434	0.270	0.536	338569	0.604	0.426
42	013	Blair County	130542	130430	-0.086	0.532	131077	0.408	0.448
42	017	Bucks County	541174	545735	0.836	0.726	537873	-0.614	0.634
42	019	Butler County	152013	153223	0.790	0.660	152898	0.579	0.635
42	021	Cambria County	163029	162949	-0.049	0.556	163876	0.517	0.481
42	027	Centre County	123786	124397	0.491	0.570	125635	1.472	0.733
42	029	Chester County	376396	380542	1.090	0.704	377088	0.184	0.535
42	041	Cumberland County	195257	195365	0.055	0.575	195256	-0.001	0.547
42	043	Dauphin County	237813	241035	1.337	0.552	239154	0.561	0.577

42	045	Delaware County	547651	554003	1.147	0.694	545064	-0.475	0.771
42	049	Erie County	275572	276888	0.475	0.529	277235	0.600	0.428
42	051	Fayette County	145351	145958	0.416	0.742	146681	0.907	0.808
42	055	Franklin County	121082	122079	0.817	0.632	122180	0.899	0.729
42	069	Lackawanna County	219039	218814	-0.103	0.532	217294	-0.803	0.732
42	071	Lancaster County	422822	423976	0.272	0.564	426528	0.869	0.523
42	075	Lebanon County	113744	113779	0.031	0.543	114518	0.676	0.589
42	077	Lehigh County	291130	291961	0.285	0.515	289980	-0.396	0.661
42	079	Luzerne County	328149	327768	-0.116	0.546	326439	-0.524	0.593
42	081	Lycoming County	118710	118822	0.094	0.538	119511	0.670	0.493
42	085	Mercer County	121003	121190	0.154	0.552	121627	0.513	0.486
42	091	Montgomery County	678111	683019	0.719	0.697	673620	-0.667	0.671
42	095	Northampton County	247105	247686	0.235	0.527	246917	-0.076	0.572
42	101	Philadelphia County	1585577	1606249	1.287	0.609	1608942	1.452	0.742
42	107	Schuylkill County	152585	153416	0.542	0.631	152989	0.264	0.525
42	125	Washington County	204584	205463	0.428	0.738	204548	-0.018	0.506
42	129	Westmoreland County	370321	371539	0.328	0.750	369009	-0.356	0.551
42	133	York County	339574	340569	0.292	0.572	341321	0.512	0.472
44	003	Kent County	161135	161498	0.225	0.654	159355	-1.117	0.776
44	007	Providence County	596270	597016	0.125	0.580	597960	0.283	0.697
44	009	Washington County	110006	110452	0.404	0.638	110982	0.880	0.633
45	003	Aiken County	120940	124770	3.070	0.542	123291	1.907	0.403
45	007	Anderson County	145196	149574	2.927	0.502	147268	1.407	0.373
45	015	Berkley County	128776	133468	3.515	0.555	132081	2.502	0.472
45	019	Charleston County	295039	304829	3.212	0.437	302751	2.547	0.580
45	041	Florence County	114344	118062	3.149	0.453	116745	2.056	0.454
45	045	Greenville County	320167	330290	3.065	0.494	325537	1.650	0.467
45	051	Horry County	144053	147841	2.562	0.452	146650	1.771	0.455
45	063	Lexington County	167611	173083	3.162	0.583	170341	1.602	0.375
45	079	Richland County	285720	295225	3.220	0.421	293299	2.584	0.564
45	083	Spartanburg County	226800	233790	2.990	0.489	230614	1.654	0.374
45	085	Sumter County	102637	105121	2.363	0.403	105017	2.267	0.500
45	091	York County	131497	133960	1.839	0.454	133717	1.660	0.409
46	099	Minnehaha County	123809	126103	1.819	0.578	124220	0.331	0.442
47	037	Davidson County	510784	532433	4.066	0.521	522044	2.157	0.617
47	065	Hamilton County	285536	293917	2.852	0.442	290664	1.764	0.512
47	093	Knox County	335749	345081	2.704	0.466	341481	1.679	0.502
47	125	Montgomery County	100498	104034	3.399	0.463	102468	1.923	0.518
47	149	Rutherford County	118570	122462	3.178	0.466	120716	1.778	0.511

47	157	Shelby County	826330	861616	4.095	0.432	847848	2.538	0.589
47	163	Sullivan County	143596	146794	2.179	0.489	145270	1.152	0.437
47	165	Sumner County	103281	105733	2.319	0.586	104756	1.408	0.343
48	027	Bell County	191088	197377	3.186	0.387	195808	2.410	0.563
48	029	Bexar County	1185394	1220995	2.916	0.498	1230141	3.638	0.744
48	039	Brazoria County	191707	196965	2.670	0.484	195577	1.979	0.374
48	041	Brazos County	121862	126396	3.587	0.520	125880	3.192	0.903
48	061	Cameron County	260120	269903	3.625	0.754	268659	3.178	0.983
48	085	Collin County	264036	271624	2.794	0.479	269149	1.900	0.412
48	113	Dallas County	1852810	1929504	3.975	0.408	1912100	3.101	0.620
48	121	Denton County	273525	282791	3.277	0.444	279483	2.132	0.495
48	135	Ector County	118934	122783	3.135	0.461	121298	1.949	0.583
48	141	El Paso County	591610	611278	3.218	0.611	617397	4.177	0.898
48	157	Fort Bend County	225421	233251	3.357	0.459	230752	2.310	0.338
48	167	Galveston County	217399	223599	2.773	0.388	221787	1.979	0.488
48	183	Gregg County	104948	107799	2.645	0.417	106936	1.860	0.522
48	201	Harris County	2818199	2939388	4.123	0.421	2915587	3.340	0.634
48	215	Hidalgo County	383545	399356	3.959	0.883	399991	4.112	0.841
48	245	Jefferson County	239397	246592	2.918	0.408	243776	1.796	0.441
48	303	Lubbock County	222636	229852	3.139	0.466	228182	2.430	0.599
48	309	McLennan County	189123	194533	2.781	0.393	193347	2.185	0.541
48	329	Midland County	106611	109988	3.070	0.466	108645	1.872	0.498
48	339	Montgomery County	182201	186761	2.442	0.500	185687	1.877	0.441
48	355	Nueces County	291145	299681	2.848	0.533	301959	3.581	0.714
48	423	Smith County	151309	155316	2.580	0.390	154321	1.952	0.391
48	439	Tarrant County	1170103	1212831	3.523	0.405	1200703	2.549	0.540
48	441	Taylor County	119655	123143	2.833	0.479	122112	2.012	0.577
48	453	Travis County	576407	594107	2.979	0.447	596444	3.360	0.663
48	479	Webb County	133239	138180	3.576	0.771	137203	2.889	1.239
48	485	Wichita County	122378	125621	2.582	0.440	124508	1.711	0.552
48	491	Williamson County	139551	143640	2.847	0.503	142663	2.182	0.376
49	011	Davis County	187941	190520	1.354	0.734	190068	1.119	0.708
49	035	Salt Lake County	725956	736793	1.471	0.635	735135	1.249	0.689
49	049	Utah County	263590	268891	1.971	0.628	271102	2.771	0.691
49	057	Weber County	158330	160566	1.393	0.581	160318	1.240	0.573
50	007	Chittenden County	131761	132031	0.205	0.587	132975	0.913	0.564
51	013	Arlington County	170936	178147	4.048	0.491	175566	2.637	0.724
51	041	Chesterfield County	209274	216590	3.378	0.584	212658	1.591	0.432
51	059	Fairfax County	818584	826402	0.946	0.575	833668	1.809	0.501

51	087	Henrico County	217881	224759	3.060	0.546	221878	1.801	0.506
51	153	Prince William	215686	218414	1.249	0.585	220359	2.121	0.425
51	510	Alexandria City	111183	112748	1.388	0.541	114451	2.856	0.771
51	550	Chesapeake City	151976	153512	1.001	0.556	155185	2.068	0.509
51	650	Hampton City	133793	139284	3.942	0.459	137415	2.636	0.617
51	700	Newport News City	170045	178053	4.498	0.468	175121	2.899	0.689
51	710	Norfolk City	261229	273457	4.472	0.444	269011	2.893	0.733
51	740	Portsmouth City	103907	108477	4.213	0.474	106837	2.742	0.695
51	760	Richmond City	203056	209959	3.288	0.549	208987	2.838	0.817
51	810	Virginia Beach City	393069	408213	3.710	0.487	402092	2.244	0.558
53	005	Benton County	112560	115161	2.259	0.556	115073	2.184	0.445
53	011	Clark County	238053	245741	3.129	0.555	241186	1.299	0.533
53	033	King County	1507319	1536441	1.895	0.519	1531673	1.590	0.612
53	035	Kitsap County	189731	196029	3.213	0.531	193702	2.050	0.425
53	053	Pierce County	586203	607187	3.456	0.502	597344	1.865	0.541
53	061	Snohomish County	465642	470715	1.078	0.625	471683	1.281	0.537
53	063	Spokane County	361364	370081	2.355	0.539	365976	1.260	0.577
53	067	Thurston County	161238	166421	3.114	0.542	164464	1.962	0.425
53	073	Whatcom County	127780	131437	2.782	0.532	130903	2.386	0.487
53	077	Takima County	188823	196444	3.880	0.499	195170	3.252	0.557
54	039	Kanawha County	207619	213488	2.749	0.492	210468	1.354	0.443
55	009	Brown County	194594	197594	1.518	0.540	195417	0.421	0.428
55	025	Dane County	367085	373810	1.799	0.541	370065	0.805	0.441
55	059	Kenosha County	128181	130580	1.837	0.548	128869	0.534	0.392
55	073	Marathon County	115400	116699	1.113	0.555	115646	0.213	0.516
55	079	Milwaukee County	959275	969329	1.037	0.459	975296	1.643	0.590
55	087	Outagamie County	140510	142519	1.410	0.543	141059	0.390	0.428
55	101	Racine County	175034	178398	1.886	0.522	176209	0.667	0.366
55	105	Rock County	139510	141935	1.709	0.558	140129	0.441	0.395
55	117	Sheboygan County	103877	105288	1.340	0.537	104218	0.327	0.445
55	133	Waukesha County	304715	306312	0.521	0.454	305387	0.220	0.361
55	139	Winnebago County	140320	142464	1.505	0.549	140855	0.380	0.418

UC RT Undercount Rate as estimated from the PES.

SE(UCRt) The sampling error of the estimated undercount rate.

UNITED STATES DEPARTMENT OF COMMER
Bureau of the Census
Washington, D.C. 20233

NOV 2 5 1992

MEMORANDUM FOR CAPE Committee

From: John H. Thompson
Chief, Decennial Statistical Studies Division

Subject: Addendum to August 7, 1992 CAPE Report

Attached to this memorandum is an addendum to the August 7, 1992 CAPE report. The addendum documents the work that has transpired since the August report was issued.

Attachment

ADDITIONAL RESEARCH ON ACCURACY OF ADJUSTED VERSUS UNADJUSTED 1990 CENSUS BASE FOR USE IN INTERCENSAL ESTIMATES

ADDENDUM TO REPORT OF THE
COMMITTEE ON ADJUSTMENT OF POSTCENSAL ESTIMATES
BUREAU OF THE CENSUS
DEPARTMENT OF COMMERCE
NOVEMBER 25, 1992

The purpose of this addendum is to summarize and document additional research conducted to examine the accuracy of a potential adjustment to the 1990 census base for use in producing intercensal estimates. The August 7, 1992 report of the CAPE, and the subsequent discussions documented in the meeting notes describe a number of areas where the committee felt more information would be helpful to the decision process. The decision to extend the period of outside comment has enabled some additional research to be carried out to more thoroughly explore a subset of these questions. This addendum summarizes that additional research.

The research has been in three basic areas -- additional analysis of accuracy based on loss functions, additional study of homogeneity within post strata, and additional work based on demographic analysis.

1. ADDITIONAL RESEARCH BASED ON LOSS FUNCTION ANALYSIS

The additional research on loss function analysis has fallen into three basic areas. First, loss functions were computed to study the accuracy of the 27 cities larger than the smallest state, Wyoming. Secondly, loss functions were computed to compare additional distributions of population shares. Finally, we computed loss functions to study some of the properties of a composite estimator suggested at the September 4, 1992 CAPE meeting.

It must be noted that the committee had significant concerns regarding the construction of the target populations which serve as the standard in loss function analysis to assess the accuracy of the adjusted and unadjusted census data. The research described below has not addressed this concern.

1.1. Cities Larger Than Wyoming

A loss function analysis was conducted to study the accuracy of adjustment for cities larger than the state of Wyoming.[1] The analysis was conducted to study the accuracy of an adjustment on the distribution of proportionate shares for just the 27 cities, for the distribution of shares for the 27 cities and the balance of the United States, and a state-by-state study of the within state population shares of the cities within the state larger than Wyoming and the state balances. This research was motivated to assess the accuracy of a suggested adjustment process that would have adjusted states and only cities larger in size than the smallest state. The following results were observed:

> The hypothesis tests at the 10-percent significance level did not indicate an improvement from adjustment to the distribution of the population shares among the 27 cities. The committee discussed this result and noted that most of these cities had high undercounts relative to the national average, but there was not a large degree of difference in the undercount among the cities.
>
> The hypothesis tests indicated that the distribution of the shares for the 27 cities and the balance of the United States was improved. Here, the committee noted that there was a difference between the undercount for the large cities, and the combined balance of the United States.
>
> The state-by-state comparisons of the within state distribution of population shares for the large cities and the corresponding state balances was mixed. For example, for the states of New York and Massachusetts the hypothesis tests indicated an improvement in the population shares between the big city and the balance of the state. These test results were observed for each of the methods of computing the target values and each of the methods of computing the loss functions. For the remaining states with one of the 27 large cities, significance at the 10 percent level was not consistently observed for each method.

---

[1]This work is documented in detail in the CAPE minutes 9-1-92, Attachment 1. Details of loss function analysis appear in "Loss Function Analysis for the Post Census Review (PCR) Estimates," Mary Mulry, 7-2-92. Cities larger than Wyoming were selected because of concerns about only adjusting states when these cities had comparable reliability.

1.2. Additional Distributions of Population Shares[2]

One criticism of the loss function analysis has been that we had been restricting our examination to the distribution of population shares for entities within specific size categories (e.g., places with population of 100,000 or more) rather than computing the loss function analysis on a distribution which includes all places or counties. We have addressed this concern by computing loss functions and associated hypothesis tests for three additional distributions of population shares:

(1) All counties;

(2) All places with 100,000 or more population and the 50 state balances of areas not included in a place with 100,000 or more population;

(3) All places with 25,000 or more population, and the balances of counties not in a place with 25,000 or more population.

Each of these three distributions completely partition the entire population of the United States. These results are discussed in section 1.4, below.

1.3. Raked Composite Estimator

Another criticism was that the composite estimation (option 4)[3] depressed the effect of adjustment among demographic groups at the national level. Therefore, a composite estimation methodology based on controlling the "50-50" estimator to national controls obtained from the Post-Enumeration Survey (PES) was also studied.[4] Eight Race/Hispanic Origin categories crossed with tenure were used as controls:

Non-Hispanic White and Other by Owner and Non-owner (2)

Black by Owner and Non-owner (2)

Non-Black Hispanic by Owner and Non-owner (2)

---

[2] A description of the details of this research is in "Additional Loss Function Analysis," John Thompson, November 4, 1992 Memorandum For CAPE Committee.

[3] This option is discussed in the August 10, 1992 Federal Register.

[4] This estimation technique is described in "Additional Loss Function Analysis," John Thompson, November 4, 1992 Memorandum for CAPE Committee.

Asian and Pacific Islanders (1)

American Indians on Reservations (1)

The controlling was carried out within each of the above categories by first calculating the difference between the full adjustment and the original 50-50 composite estimator at the national level. This difference was then allocated sub-nationally within the control categories using the proportional distribution of the original 50-50 composite estimator. Statistically, this follows a technique referred to as "raking", leading to the terminology of Raked Composite Estimator. These results are discussed below.

1.4. Summary of Results

The results of the loss function analysis for the measures described in sections 1.2 and 1.3 are summarized in the form of significance values for the hypothesis testing. We have restricted our analysis to two of the methods of calculating the target populations -- PROPUC and GROSDSE without correlation bias. (These methods of computing the targets are described in "Total Error for Postcensus Review Estimates of Population" by Mary Mulry, July 7, 1992). We were not able to carry out the analysis for all of the methods of computing the target populations. We selected these two methods for study because we believe that they will cover the range of alternative methods to calculate the target populations. We also excluded the correlation bias modeling because there were still many questions about how to estimate correlation bias and how to adequately allocate the estimate of correlation bias to all geographic areas of interest. The effect of not including correlation will also tend to be conservative, since including measures of correlation bias would most likely favor adjustment.

1.4.1 Weighted Squared Error

This section summarizes the CAPE presentation and discussion of these results. The committee discussion was centered on weighted squared error results. Table 1 presents the significance values for the weighted squared error loss function hypothesis test results.[5] We show data for the full adjustment, the raked composite, and the 50-50 composite. The results are displayed for states, counties, and places. We show the results for the previous size category distributions, and for the three new distributions. A summary of key results follows:

[5]These results are discussed in more detail in "Additional Loss Function Analysis," John Thompson, November 4, 1992 Memorandum for CAPE Committee.

(1) The hypothesis test significance values for the full adjustment indicate little evidence of an improvement from adjustment for the PROPUC target population method for most size categories -- they are well above 0.10. However, for the loss functions reflecting places of 100,000 or more population and state balances (the last line in Table 1), the significance level approaches 0.10.

(2) The hypothesis tests for the GROSDSE target population method are much more significant than for the PROPUC method, thus indicating more evidence for improvement due to adjustment. This is particularly the case for the new size categories for counties and places.

(3) The hypothesis tests for the raked composite are much more significant than for the full adjustment, indicating more evidence for improvement. These tests are significant at the 10-percent level for the new size categories, for both target population methodologies.

(4) The significance values for the 50-50 composite (where available) are similar to the raked composite estimator.

(5) The hypothesis tests for places between 50,000 and 99,999 population are more significant than the tests for areas with 100,000 or more population.

1.4.2 Squared Error and Relative Squared Error

Tables 2 and 3 give significance values for loss functions based on squared error and relative squared error, respectively.

1.5. Summary of Committee Discussion

The committee discussed these data and noted that while gains had been achieved in reducing sampling error, the raked composite estimator depended more heavily on the assumption of homogeneity (discussed in more detail below). Many on the committee expressed concern with balancing the reduction in sampling error with the greater dependence on the homogeneity assumption. Given these concerns, there was general consensus that the raked composite estimator offered great potential for future research. However, there was not currently enough information available to select this estimator as superior to the full adjustment which had been more thoroughly studied and discussed.

The committee noted that there was some evidence that large areas (greater than 100,000 population) were improved by adjustment when compared to the balance of state. However, the committee also noted that these loss function results should be treated with caution, since they were subject to the same limitations as noted in the August 7, 1992 committee report.

2. ADDITIONAL RESEARCH ON THE HOMOGENEITY ASSUMPTION

The validity of the homogeneity assumption was one of five basic issues addressed by the CAPE report of August 7, 1992. The report summarized the status of knowledge at that date by the following (p. 25):

> Summary: Just as in July 1991, the results on whether the homogeneity assumption holds are inconclusive. The new research used to examine the homogeneity assumption (called artificial population analysis) indicates that the assumption does not hold when the bias in the estimate gets to be about 25% or higher. Since the bias in the Post Enumeration Survey (PES) estimate as currently measured is 22% to 45%, the Committee was concerned.

New analysis has refined the use of the artificial population analysis to examine quantitatively the effect of departures from the homogeneity assumption and to assess the performance of the loss function analysis in the presence of heterogeneity. Reexamination of the evidence has identified areas of incompleteness in the analysis of the previous findings about the effect of bias on the loss function analysis.

2.1. Refinements to the Analysis of the Artificial Populations

Much of the previous research of the artificial populations focused on assessment of whether, in the absence of sampling variance and bias, the 357-post-strata estimator would represent an improvement in the true distribution over the census distribution. This research left largely unanswered questions about the possible size of the effect that departures from the homogeneity (or synthetic) assumption could have, and how such departures would interact with other aspects of the PES analysis, especially the loss function analysis. The results of the reanalysis described below were presented to the committee on November 5, 1992. The committee has not conducted an extensive discussion of these new findings.

The reanalysis focused on three measures: squared error, weighted squared error, and relative squared error. (Measures based on absolute error appeared to present difficult technical issues and were not considered.) The analysis was restricted to the state level. A first part of the analysis addressed the question:

> Q1: Compared to other errors in the PES estimation, how much effect could departures from the homogeneity assumption have on the errors of the PES estimates?

This question was addressed by reexpressing previous findings for the artificial populations by forming the ratio of losses under the adjustment compared to no adjustment. Thus, a ratio of 0 would indicate that the homogeneity assumption was completely satisfied, 0.20 indicates that the PES estimator could potentially capture 80 percent of the underlying variation in the corresponding artificial population, a ratio of .80 indicates that adjustment would capture only 20 percent of the underlying variation. Although ratios above 1.0 are theoretically possible, none were observed. The results are included in the minutes of the CAPE for November 5, 1992. Although ratios ran a wide gambit, going down as far as a highly favorable .11, 12 of the 24 ratios exceeded .50. Such evidence indicates a strong possibility that the 357 post-strata design may capture only about half of the true state-to-state variation in undercount.

The strong possibility that errors due to heterogeneity could be as large as half the errors in the census now appears consistent with the observation made by one reviewer that the errors due to heterogeneity could be larger than all of the errors in the PES accounted for by the total error model.

Given this potentially high level of error, it became critical to assess how heterogeneity would affect the loss function analysis. (If heterogeneity was found conclusively to be quite small, then it could be successfully argued that heterogeneity could only have a small impact on the validity of the loss function analysis.) On the other hand, the reanalysis was still consistent with earlier findings, namely, that the bias due to heterogeneity does not, by itself, obviate the ability of the adjustment to make improvements on the census.

The second question is therefore:

> Q2: How does heterogeneity affect the rest of the PES analysis? In particular, in the presence of heterogeneity, can the PES loss function analysis still reliably measure the improvement, if any, that adjustment makes?

The artificial populations were also used to assess this second question. Since the loss function analysis compares the PES estimates and the census to target values constructed through the synthetic estimator used in the PES, the artificial populations can be used to ascertain whether comparison to such targets correctly states, understates, or overstates the actual improvements of adjustment, which are determined by comparing the census and adjusted distribution to the true census values.

Largely, the evidence supported the continued use of the loss function analysis as a measure of the net improvement, although with qualifications. In particular:

- For the majority of populations, the loss function analysis was actually conservative, tending to understate the true improvement in distribution by using target populations constructed from the synthetic model, compared to the actual advantage of adjustment over the census when the true state values were used as a standard for comparison.

- For two populations, poverty and mobility, the balance between the loss function analysis and the actual improvement appeared about right, in some cases overstating the advantage of adjustment slightly.

- In one instance, the artificial population based on unemployment, the synthetic model was the least successful, explaining only about 20 percent of the variability at the state level. Furthermore, the loss function analysis was seriously distorted, presenting a seriously misleading measure of the improvement due to adjustment.

Some attempts had been made to assess the interaction of sampling error on the analysis by assigning sampling variance to the post-strata. Thompson, and Alberti, discussed some of these findings in a memorandum to the CAPE.[6] Their findings indicated that sampling variance would raise serious questions against adjustment. However, the results of the analysis depended on how much variance was assigned. Thompson and Alberti did not have direct estimates of variance for the artificial population variables. In place of arbitrary decisions about variance, Fay, in recent work (memorandum of Nov. 18, 1992 to the CAPE Committee) calculated sample estimates for the 357 post-strata for 5 of the 8 artificial populations, based on the PES sample blocks only, with the appropriate survey weights. These results have not been discussed by the committee. Several members of the committee view them as being more supportive of adjustment, but questions still remain regarding how much variance must be assigned.

2.2. Reexamination of Bias with the Artificial Population Analysis

The previous CAPE report asserted that the artificial population analysis had shown that the improvement from adjustment apparently vanished when the PES estimates were subject to biases on the order of 25 percent, as noted in the cited summary. In fact, reexamination of the findings presented to the committee revealed that the results were different from the interpretation given them in the earlier report.

---

[6]"Additional Results for Artificial Populations'" John Thompson, September 2, 1992 Memorandum for CAPE Committee.

Although most of the CAPE analysis focused on distributive accuracy, the statistical analyses leading to the figure of 25 percent bias were all based on numeric accuracy. Initially, it was thought that the method of modeling the bias would have no effect on the loss functions for population shares. Further analysis has indicated that this is not the case, and that the loss function analysis for population shares is more robust to our method of modeling bias. This work was not available to the committee for discussions regarding the failure of the homogeneity assumption. A more detailed analysis of this work combined with alternative methods of modeling bias should be carried out in future studies to learn more about the effect of bias on the loss function analysis for population shares.

3. ADDITIONAL RESEARCH ON THE CONSISTENCY OF PES ESTIMATES OF COVERAGE WITH DEMOGRAPHIC ANALYSIS AND OTHER INDICATORS OF COVERAGE FOR SUB-STATE AREAS

The CAPE discussed the consistency of the PES and demographic analysis estimates in the August 7, 1992 report. At that point, the committee generally felt that the PES estimates met their face validity expectations at the State level with some individual state exceptions. Since August, additional research has been conducted to examine the face validity of the PES estimates for large sub-state areas based on demographic indicators.

Direct demographic estimates of the population under age 10 were produced and compared to the PES estimates. This work was accomplished in 40 states for 132 large counties and state balances (total of 172 individual areas). Additional work was also carried out for proxy measures of coverage at the sub-state level. Measures such as percent minority, percent renter, substitution rates, and poverty rates were used.

These results were briefly discussed at the November 5, 1992 CAPE meeting.[7] A detailed discussion of these results will be documented in a future internal memorandum from the Population Division. The results tend to indicate that there are very general patterns of agreement between the PES and demographic analysis results. There has been no extensive review or discussion of these findings by the committee, therefore, no conclusions can be stated.

---

[7]This discussion does not appear in the notes of the November 5, 1992 meeting. The results were merely mentioned in passing.

4. FINAL SUMMARY

The additional research described above has addressed some of the concerns documented in the initial report of the committee August 7, 1992. The general conclusions from that report remain much the same.

4.1 The August 7, 1992 report indicates that the committee concluded that on average, an adjusted state base would be more accurate than an unadjusted state base for use in intercensal estimates. This is still the case. The research based on loss functions since August 7, 1992 has indicated that additional evidence exists that adjustment will improve the distribution of population shares for large places (100,000 or more population) compared to the balance of state.

4.2 The research on the homogeneity assumption has indicated that the total error model does not include a complete measure of the error due to failure of the homogeneity assumption. The research also indicated that the loss function analysis based on the total error model was somewhat robust to this problem, and could be viewed as a measure of net improvement. The research also indicated that more information should be gathered regarding the effect of measurement biases on homogeneity.

Table 1 Significance Probabilities for Loss Functions Analysis for the Weighted Squared Error Loss Function

	Full Adjustment		Raked Composite		50:50 Composite	
	PROPUC	GRODSE	PROPUC	GRODSE	PROPUC	GRODSE
STATES						
All States	0.29	0.13	0.13	0.04	0.12	0.05
COUNTIES						
All Counties	0.29	0.12	0.09	0.02	NA	NA
Less than 200K	0.89*	0.61	NA	NA	NA	NA
200K or More	0.06*	0.08	0.04	0.02	0.04	0.02
PLACES						
25K or More	0.27	0.20	0.08	0.05	NA	NA
25K+ County Bal	0.27	0.10	0.08	0.04	NA	NA
50K-100K	0.25*	0.33	0.21	0.10	0.14	0.06
100K or More	0.22*	0.52	0.26	0.25	0.25	0.23
100K+ State Bal	0.14	0.03	0.04	0.01	NA	NA

NOTES:

- PROPUC is the PROPUC without correlation bias target except where indicated otherwise
- GRODSE is the GRODSE without correlation bias target
- An '*' indicates that the PROPUC with correlation bias value is given since the PROPUC without correlation bias value is not available

Table 2 Significance Probabilities for Loss Functions Analysis for the Squared Error Loss Function

	Full Adjustment		Raked Composite		50:50 Composite	
	PROPUC	GRODSE	PROPUC	GRODSE	PROPUC	GRODSE
STATES						
All States	0.21	0.08	0.10	0.03	0.09	0.04
COUNTIES						
All Counties	0.09	0.05	0.04	0.02	NA	NA
Less than 200K	0.84*	0.56	NA	NA	NA	NA
200K or More	0.05*	0.09	0.05	0.03	0.04	0.03
PLACES						
25K or More	0.46	0.49	0.31	0.33	NA	NA
25K+ County Bal	0.14	0.13	0.09	0.10	NA	NA
50K-100K	0.24*	0.32	0.20	0.09	0.13	0.06
100K or More	0.45*	0.81	0.57	0.64	0.57	0.60
100K+ State Bal	0.19	0.04	0.06	0.01	NA	NA

NOTES:

- PROPUC is the PROPUC without correlation bias target except where indicated otherwise
- GRODSE is the GRODSE without correlation bias target
- An '*' indicates that the PROPUC with correlation bias value is given since the PROPUC without correlation bias value is not available

Table 3 Significance Probabilities for Loss Functions Analysis for the Relative Squared Error Loss Function

	Full Adjustment		Raked Composite		50:50 Composite	
	PROPUC	GRODSE	PROPUC	GRODSE	PROPUC	GRODSE
STATES						
All States	0.55	0.26	0.27	0.11	0.26	0.11
COUNTIES						
All Counties	0.90	0.56	0.71	0.33	NA	NA
Less than 200K	0.93*	0.63	NA	NA	NA	NA
200K or More	0.09*	0.10	0.05	0.02	0.06	0.03
PLACES						
25K or More	0.21	0.09	0.05	0.01	NA	NA
25K+ County Bal	0.83	0.57	0.60	0.35	NA	NA
50K-100K	0.25*	0.35	0.22	0.11	0.14	0.07
100K or More	0.13*	0.26	0.15	0.09	0.13	0.09
100K+ State Bal	0.19	0.10	0.06	0.02	NA	NA

NOTES:

- PROPUC is the PROPUC without correlation bias target except where indicated otherwise
- GRODSE is the GRODSE without correlation bias target
- An '*' indicates that the PROPUC with correlation bias value is given since the PROPUC without correlation bias value is not available

11/ 24/92

COMMITTEE ON ADJUSTMENT OF POSTCENSAL ESTIMATES (CAPE)
DISTRIBUTION LIST

B. Bryant	DIR	2049-3
H. Scarr	DIR	2049-3
P. Bounpane	DIR	3049-3
C. Jones	DIR	2031-3
R. Tortora	DIR	2031-3
W. Butz	DIR	2061-3
R. Fay	DIR	3067-3
P. Schneider	FLD	3023-3
D. Blass	FLD	3027-3
M. Mulry	2KS	3525-3
J. Thompson	DSSD	3785-3
J. Long	POP	2019-3
G. Robinson	POP	2022-3
M. Batutis	POP	2336-3
H. Hogan	SRD	3207-4
D. Whitford	DMD	3554-3
T. Walsh	DMD	3586-3
C. Miller	PPDO	2430-3
L. Lyons	PPDO	1085-3

Please direct any changes to Lana Sylier 763-5337.

11/ 24/92

COMMITTEE ON ADJUSTMENT OF POSTCENSAL ESTIMATES (CAPE)
DISTRIBUTION LIST

B. Bryant	DIR	2049-3
H. Scarr	DIR	2049-3
P. Bounpane	DIR	3049-3
C. Jones	DIR	2031-3
R. Tortora	DIR	2031-3
W. Butz	DIR	2061-3
R. Fay	DIR	3067-3
P. Schneider	FLD	3023-3
D. Blass	FLD	3027-3
M. Mulry	2KS	3525-3
J. Thompson	DSSD	3785-3
J. Long	POP	2019-3
G. Robinson	POP	2022-3
M. Batutis	POP	2336-3
H. Hogan	SRD	3207-4
D. Whitford	DMD	3554-3
T. Walsh	DMD	3586-3
C. Miller	PPDO	2430-3
L. Lyons	PPDO	1085-3

Please direct any changes to Lana Sylier 763-5337.

UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, D.C. 20233

NOV 2 4 199_

MEMORANDUM FOR CAPE Committee

From: John H. Thompson
Chief, Decennial Statistical Studies Division

Subject: Appendix A -- Discussion of Technical Comments

Attached to this memorandum is Appendix A -- discussion of technical comments received from outside reviewers.

Attachment

November 24, 1992

Appendix A: Discussion of Technical Issues Raised by Outside Comment

A number of important technical issues have been raised from the public commentary. Most of these issues have been included in the CAPE discussions and documented in the report of the committee or in the meeting notes.

The majority of the technical comments raised issues against adjustment of the 1990 census base. The concerns expressed were for the most part related to the analysis conducted by the Census Bureau, and the assumptions that underlie the analysis, and the PES estimation. The opinions expressed in support of adjustment generally recommended that the full adjustment be carried out for all levels of geography. A more detailed discussion of these issues is given below:

1. Homogeneity

Several writers pointed out the critical nature of the dependence of the adjustment on the homogeneity assumption. This assumption states that the undercount rates should remain fixed within each of the 357 post-strata. Although post-strata typically fall in several states and numerous counties, within any one post-stratum, undercount rates are assumed to remain fixed from one state to another, from one county to another, etc. The undercount rate is assumed to vary from one post-stratum to another, however; indeed, this variation is the basis of the adjustment. To the extent that the homogeneity assumption is violated, the undercount rates are said to be "heterogeneous." By adjusting all persons in a post-stratum by the same factor, the PES estimator assumes that the homogeneity assumption holds.

When homogeneity is defined in this manner, it is virtually self-evident that the assumption cannot hold exactly. The key issue, from the perspective of CAPE, is whether the homogeneity assumption represents an adequate approximation to the distribution of undercount to result in an improvement in the overall distribution of population totals and shares.

The committee agrees that this assumption is one of the more vulnerable aspects of the PES design. Discussions and investigations of the homogeneity assumption occurred in the Undercount Steering Committee's deliberations prior to July 15, 1991, and the issue has been one of ongoing interest for the CAPE. Additional research has followed the report of the CAPE, inspired, in part, by comments received during the period of public comment.

Some reviewers provided only general remarks about their concerns with the homogeneity assumption, but others provided specific insights that spurred further investigation. One created a U.S. map showing the high degree of association between the adjustment at the state level and the groupings of states into the 4 census regions, North East, South, Midwest, and West, used in defining many of the post-strata. The reviewer showed maps of other characteristics, such as poverty rates, which do not exhibit so marked a regional character as the adjustments. Researchers at the Census Bureau subsequently reexamined the series of characteristics employed in defining the 8 sets of "artificial populations" - simulations of characteristics based on census data, such as the poverty rate, in a similar manner. To varying degrees, the Census Bureau's investigations confirm the point made by this reviewer, that is, that the adjustment methodology tends to emphasize regional aspects of the characteristic being estimated while missing or understating other components of state-to-state variation.

Another reviewer provided calculations showing that it was possible that departures from the homogeneity assumption, that is, heterogeneity, might account for more error in the PES adjustments of states than all the components of error estimated and included in the Census Bureau's total error model. This reviewer appeared to argue that a decision to adjust could not be reliably made when such a potentially large component of error had not been incorporated.

Consequently, research in this area has continued during the fall. The principal part of this research employed "artificial populations" based on actual population characteristics measured from the census, which were discussed in the previous report of the CAPE. Results included in the CAPE minutes of November 5, 1992 showed that, at the state level, the effect of departures from the homogeneity assumption tended generally to be large for the artificial populations. In fact, these investigations supported the strong possibility that the error due to heterogeneity could indeed be larger than all other sources of error in the adjustment, as one reviewer had suggested. In turn, the loss function analysis tended to understate the errors from adjustment, because heterogeneity bias tended to add to errors that the loss function analysis estimated for adjustment.

On the other hand, the investigations continued also to support the premise that the PES adjustment could still, on average, make improvements to the overall population shares. When heterogeneity bias is present, the results for artificial populations showed that the loss function analysis would tend to understate the errors of both the adjustment and of non-adjustment. If the loss function analysis understated the errors of adjustment and non-adjustment by equal amounts, then its estimate of the net difference would be correct. In fact, the analyses showed that, for a majority of the 8 populations, the loss function analysis would approximately correctly indicate or understate the net advantage from adjustment. There was an exception: the results for the artificial population based on unemployment gave unacceptable results, that is, the resulting loss function analysis would appear to have exaggerated substantially the net gains from adjustment.

2. Insufficient Sample Size

Some writers argued that the PES sample size was insufficient to permit an adjustment. These reviewers based their conclusion on comparisons to sizes of other samples with which they were familiar. These arguments were not reinforced, however, by explicit calculations showing in what sense the sample size was too small.

The issue of sample size is linked directly to the level of sampling variance, since increasing sample sizes predictably reduces sampling variance while not reducing most components of nonsampling error. The Census Bureau's total error model and loss function analysis were specifically designed to test whether the sample size was sufficiently large to obtain an improvement in the estimated total population and shares. In general, there is not one specific sample size that can be said to be large enough, since whether improvements can be made depends also on how much the undercount varies from one state to another, levels of measured nonsampling error, and the estimation procedure. The use of the 357-post-strata design reduced the effect of sampling variability considerably. By weighing the advantage that the PES would appear to accrue in correcting the census for large errors in states such as California against the small errors that would occur in estimating other states close to average undercount, the loss function analysis indicated that the PES sample size was sufficient to control uncertainty from sampling.

3. Post-Stratification

The Census Bureau decided to use a revised post-stratification scheme to control sampling variability instead of using a smoothing model. Several comments were received applauding this decision. However, some of these reviewers claimed that the post-stratification was data-driven. The end result of this was that the estimates of sampling error would be too low therefore causing the loss function analysis to unduly favor the adjustment.

One reviewer indicated that the revised post-stratification was acceptable, but indicated that a smoothing model would have been preferable. In addition, this reviewer indicated that an alternative technique to control sampling variability would have been to collapse the original post-stratification scheme based on

1392 categories. This would have had the effect of retaining greater homogeneity within post-strata. In the end, however, the reviewer (and the committee as well) felt that the revised 357 post-stratification scheme was superior to no adjustment.

These issues were discussed at various points by the committee. The committee was almost unanimous in deciding that smoothing would not be used in producing the revised post-stratified estimates. The committee was also pleased with the resulting post-stratification scheme. The committee recognized the danger of post-stratification, after data had been examined, and these concerns were documented in committee discussions.[1] This had some bearing on the general concerns that the committee expressed regarding the loss function analysis.

4. Correlation Bias

Correlation bias was widely discussed both internally and externally. Some reviewers generally noted that a correction based on correlation bias would be conservative in that it would not go far enough in correcting the differential undercount.

Other reviewers noted that at the national level there was clear evidence of correlation bias. However, they claimed that problems resulted because there were no direct measures of correlation bias sub-nationally. It was not clear to these reviewers that the methods of modeling correlation bias to produce sub-national estimates was appropriate, and there was concern that no supporting empirical evidence existed. Therefore, these reviewers were not convinced that the adjustment would improve the distribution of population shares sub-nationally.

The CAPE also expressed many of these same concerns as documented in the August 7, 1992 report, and in various meeting minutes.[2] The general conclusion of the committee was that correlation bias should be a component of total error. However, there were concerns expressed regarding the methods of estimating and allocating it. The committee requested that loss function analysis be done with and without correlation bias. Each committee member then had to make individual judgements about how to analyze the results.

---

[1]CAPE minutes from 9-18-91, 12-2-91, 12-30-91, 1-13-92, 1-27-92, 1-10-92, 1-16-92, 4-6-92, 4-22-92.

[2]CAPE minutes from 4-20-92, 6-1-92, 4-9-92, 6-11-92, 6-29-92, 9-1-92, 4-22-92

5. Total Error Model

Some reviewers viewed the total error model as being complete, and when combined with the loss function analysis supportive of an adjustment. One reviewer noted that he felt that the total error measurement of correlation bias was understated and a more accurate measurement would favor adjustment more than the current estimates.

There were other reviewers who did not believe that the total error model covered all sources of error. These reviewers cited various sources of error that they felt were omitted, such as: uncertainty from the choice of post-stratification or uncertainty from failure of the homogeneity assumption. These reviewers also felt that many of the sources of error included in the total error model were not measured accurately. They cited biases arising from imputation of missing data, fabrication error, and misreporting census day address as being particularly understated.

The CAPE discussed the total error model at great length.[3] The committee felt very confident that all components of error, except for bias due to failure of the homogeneity assumption, had been listed and considered. However, the committee could come to no agreement about the adequacy of the levels of error measured for the total error model components. The committee concluded in general to use caution in evaluating the results of the loss function analysis since the target numbers used in loss function analysis were so dependent on the levels of estimated bias.

6. Loss Function Analysis

Some reviewers viewed the loss function analysis as being very supportive of adjustment, and that the improvement indicated by the loss function analysis was an understatement (correlation bias was underestimated in the total error model).

Other reviewers generally had two major sources of concern regarding the loss function analysis: (1) There are uncertainties in the adjusted estimates that are not included in the loss function analysis, including uncertainty from failure of the homogeneity assumption, and from the choice of post-stratification. (2) There are concerns with the methods used to model the total error estimates of bias to create the target populations. In addition, one reviewer expressed concerns regarding the levels of significance reported for the loss function hypothesis tests.

---

[3]CAPE minutes from 4-20-92, 6-1-92, 6-9-92, 4-13-92, 4-22-92

The CAPE also discussed the loss function analysis in great detail.[4] In particular, the comments regarding uncertainty due to failure of the homogeneity assumption lead to some of the additional research reported in the addendum to the CAPE report. In general the committee accepted the loss function results keeping in mind a number of caveats.

7. Have All Questions Been Answered?

Many of the reviewers noted that there were many areas where additional research would provide useful information to inform the decision process. Some reviewers felt that sufficient information was available. Other reviewers generally, felt that more information was needed in order to justify an adjustment.

The August 7, 1992 report of the CAPE also indicates a number of areas where more research could be applied. Some of this research was continued and is reported in the addendum to the report. Many questions still remain. In spite of a desire for more complete information, the CAPE was able to reach general consensus that on average an adjustment to the 1990 census base would improve the intercensal estimates for states. For areas below the state level, the committee was able to reach no general consensus.

PE minutes from 4-27-92, 5-4-92, 6-9-92, 6-24-92, 6-29-92, 7-13-92, 9-1-92, 11-5-92, 4-22-92, 9-4-92

[4]CA 7-6-92,

Technical Assessment of the Accuracy of Unadjusted versus Adjusted 1990 Census Counts

Report of the Undercount Steering Committee
Bureau of the Census
Department of Commerce
June 21, 1991

RECOMMENDATION

The Undercount Steering Committee (USC) has completed its assessment and concludes that a statistical adjustment of the 1990 census leads to an improvement in the counts. A large majority of the Committee subscribes to this decision.

The Committee's assessment is based on statistical evidence from the post-enumeration survey (PES), qualitative and quantitative evaluations of the PES, comparisons to results from demographic analysis (DA), evaluations of DA itself, and professional judgments based on statistical reasoning and substantive demographic expertise. The assessment is a technical one, ignoring the political ramifications of the adjustment.

A minority disagrees with the Committee's decision because they believe that reasonably complete analyses of results have yet to be performed. However, they would support the use of adjusted counts in the intercensal estimates program, and they believe that an alternative adjustment using the PES data to adjust for post-strata with large measured undercounts might be acceptable to them. This report presents both majority and minority views.

The Committee focused on whether the adjusted counts on average are better than the unadjusted. In doing so, it acknowledges that both sets of counts have potential weaknesses--the unadjusted can suffer from systematic undercounts; the adjusted can introduce undesirable random variability or new biases into results. The unadjusted census is found to undercount the same well-identified subpopulations as previous censuses. The task of assessing the adjustment became a process of judging 1) whether it removes the systematic weaknesses of the unadjusted counts, and 2) whether it does so in a stable fashion on the national, state, and local levels.

It is impossible to be certain that adjusted counts are more accurate for each and every area in the country. The Committee could assess the quality of adjustment better for larger areas than for smaller areas. For a small number of States, there is concern that the adjusted counts are less accurate. However, the majority of the Committee concludes that the accuracy gain for the majority of States offsets the small possible losses in relatively few States. It is understood that for smaller areas, those with less than 100,000 population, proportionately more units would have less accurate adjusted counts than unadjusted. However, the majority finds that on the average the adjustment is beneficial at lower levels also.

The majority's conclusion is based in large part on their finding that the post-enumeration survey is a measurement instrument of unusually high quality. We know of no survey that has been subjected to the scrutiny given the PES. Twenty-two evaluation studies were conducted to measure error in the PES. The PES is designed to accurately measure and adjust for undercount if errors can be controlled. Although there is concern about certain findings of the studies, most supported the conclusions that the PES was very successful. Thus, the majority judges that an adjustment based on the PES would ameliorate the undercount of minority groups and improve the accuracy of counts for the Nation, States, and places of 100,000 population or more.

There is little direct evidence to judge whether adjusted counts are more accurate for places under 100,000 population. Time did not allow for full simulations of accuracy for smaller areas. There was some evidence from the loss function analysis, but there was no independent evidence with which to compare it. Even if such simulations had been done, they would have been limited by the nature of the PES as a sample survey with relatively smaller sample sizes for the PES evaluations. Even so, in the absence of direct evidence to the contrary, the majority concludes that adjusted counts are generally more accurate at lower levels. Therefore, the improvement in counts on the average for the Nation, States, and places over 100,000 population outweighs the risk that the accuracy of adjusted counts might be less for some smaller areas.

The Committee decision is based on the data available for the method chosen in order to prepare this report in time for the Director of the Census Bureau to make recommendations to the Secretary of Commerce. The Secretary must make his choice by the July 15, 1991 court-ordered deadline. The Committee thinks that with more time, a methodology would likely be found that would produce a better set of adjusted counts. However, the Committee was charged with evaluating the current methods, which were found to be acceptable. The majority expects that any adjusted counts produced by any future methodology, while different from those available today, would still show that adjusted counts are more accurate than the census. A minority disagrees with making the decision without what they consider an adequate review of data.

THE UNDERCOUNT PROBLEM AND THE PES PROGRAM

Having a census with full coverage has long been a goal of the Census Bureau. Though overall coverage has tended to get better from census to census, there has been a consistent difference in the coverage for certain minority populations when compared to the rest of the population. This coverage differential was measured consistently by both the PES and demographic analysis (DA) for the 1990 census. The Census Bureau faced the question of whether to adjust the counts with regard to the 1980 census, and could not conclude that adjusted counts were more accurate than census counts. Because of that, the Census Bureau instituted many programs in the 1980's designed to improve coverage measurement.

The Census Bureau established a goal of conducting two coverage measurement programs (PES and demographic analysis) and some 33 evaluations of these programs. Nearly a year and a half ago, the Bureau developed a plan to complete these activities in order to make the present assessment. This was the most intensive research and evaluation project the Census Bureau has ever undertaken.

We are pleased to report that this work has been completed. Completion was accomplished because the Bureau management was committed to this goal, and staff at all levels--in the field, in the processing offices, and the technical, support, and coordinating staff at headquarters--committed themselves with extraordinary dedication to complete the work.

Thanks to the completion of the coverage measurement and evaluation programs, we have more information than ever before to inform our assessment. The Undercount Steering Committee spent numerous hours in session in the last 2 months reviewing and evaluating the studies and reaching its conclusions. This was in addition to countless hours of individual review and preparation in the same time period and many months of prior examination of adjustment-related issues.

Primarily, the Undercount Steering Committee concerned itself with Guidelines 1-3, the "technical" guidelines that relate to the accuracy of adjusted counts. A discussion of the Committee's assessment, following Guidelines 1-3, follows. We also discussed Guideline 6, which relates to whether sufficient data are available and whether analysis of the data is complete enough to make our determination. We include comments on guideline 6 within the discussion of Guidelines 1-3, as well as in a separate section. In appendix 1, we list the members of the Undercount Steering Committee. In appendix 2, we provide selected tables presenting the post-enumeration survey and demographic analysis estimates. In appendix 3, we provide a brief background sketch of the coverage measurement and adjustment process for the general reader. In appendix 4, we list the PES and DA evaluation studies. In appendix 5, we provide information on the 1,392 post-strata. Appendix 6 is a more technical discussion of statistical issues relevant to the Committee's assessment.

DISCUSSION OF GUIDELINES

GUIDELINE 1

Guideline 1 states that "The Census shall be considered the most accurate count of the population of the United States, at the national, state, and local level, unless an adjusted count is shown to be more accurate. The criteria for accuracy shall follow accepted statistical practice and shall require the highest level of professional judgment from the Bureau of the Census. No statistical or inferential procedure may be used as a substitute for the Census. Such procedures may only be used as supplements to the Census."

Summary of Guideline 1.--In assessing this guideline the Undercount Steering Committee (1) compared the demographic analysis estimates and the PES estimates, (2) assessed errors in the PES, and (3) reviewed a total error model and loss function analysis as an aid to determine whether adjusted counts are overall more accurate than the census. In the discussion of each of these, the Committee also considered the consistency of the adjusted counts with well documented demographic trends for areas.

In summary, the majority is convinced by the relatively low levels of error in the post-enumeration survey as measured by the studies. They also rely on the fact that the loss function analysis shows that in a large majority of areas, adjusted counts are closer to "truth" than census counts. The majority also concludes that face validity analysis does not refute the loss function results. The minority is concerned that the PES evaluation results are not accurately reflecting total error, and therefore, they are less confident in total error/loss function analysis. The minority also is concerned that model errors are not appropriately reflected in the total error results. All members are comfortable that for some evaluation post-strata (summaries of the production post-strata), the adjusted counts are more accurate than the census.

Comparison of Demographic Analysis and Post-Enumeration Survey. Before using demographic analysis as one standard of comparison, 11 evaluation projects were conducted (see appendix 4). These provide, for the first time, measures of uncertainty about the DA estimates. DA and PES measure similar national undercount rates (1.85 and 2.07 percent, respectively). In addition, the age-sex-race patterns of undercount estimates for PES and DA are similar, with higher undercounts for Blacks than Non-Blacks and males than females. However, DA shows higher undercounts of Black males than PES. If DA is accepted as truth, then the adjusted counts will not fully correct for the Black male undercount, but will move the counts in the right direction, on average. PES shows higher undercounts of females and two age groups than does DA. If DA is accepted as truth, then the adjustment will over-correct for females and persons in those two age groups. Taking into account the comparison for all 24 age-sex-race groups, the majority conclude that the similarity between the PES and DA results is sufficient to support the judgment that the PES is reflecting real undercounts in the census. The minority places large weight on the differences between PES and DA undercount estimates.

Assessment of Errors in the PES. The Committee reviewed 22 evaluation studies designed to measure sources of error in the PES. (See appendix 4.) These studies provide quantitative estimates of the quality of various components of the PES. These include quality of matching cases between the PES and the census, quality of reporting census day address, the extent of fabrications during the data collection, the extent of missing data on individual questions, the consistency of search procedures between the P- and E-samples (see page 2 of appendix 3 for a description), correlation bias, and random error. The estimates of individual components are combined in a quantitative total error model, that summarizes the total quality of the PES, and corresponding loss functions to answer the question of whether the adjusted or unadjusted census is more accurate. The Committee first examined the individual studies one by one, seeking evidence of fatal flaws in the PES. All members would have preferred evaluation studies that produced more stable estimates, but the majority finds no evidence of fatal flaws. The minority saw the same evidence and question the adequacy of several studies. For example, they are concerned that the high PES estimate of Hispanics might be explained by matching error not measured in the matching error study.

No evaluation study could directly measure correlation bias below the national level, and so all assessments of its effect depend on judgment. Generally, correlation bias in the PES would, in the absence of other errors, lead to an understatement of undercount by the PES. The majority notes that the range of correlation bias estimates still produces adjusted counts more accurate than the unadjusted. The minority remains concerned about estimates of correlation bias that are contrary to expectations and past experience.

Overall, the minority is concerned that the evaluation studies understate the level of error, especially when compared to error estimates from past censuses and tests. The majority notes that important changes in methods were made between those efforts and the 1990 PES, and that larger levels of error are not implied by the comparison with demographic analysis estimates of undercount. All members would prefer more time to analyze the results of the studies in more detail, in order to reduce their reliance on the conclusions reached by the project managers. There is a majority conviction, however, that major flaws in the PES are not likely to be uncovered as the data are further analyzed.

Professional staff examined adjusted counts for various geographic levels and compared them with the census, postcensal estimates carried forward from the 1980 census, and demographic analysis. They prepared reports that detailed their findings of how adjusted counts compared to other estimates or, in other words, how much "face validity" the adjusted counts had. These reports generally state that adjusted counts have face validity, at least for large geographic areas. Although some unusual situations were noted, the vast majority of comparisons made at various geographic levels are favorable to the adjusted counts. The majority finds that face validity assessments generally add credence to the PES results. Face validity assessments of various data sources do, however, raise concerns among a minority about error in the PES. These include high measured undercounts in unexpected places; relatively low undercounts in some major cities in the Northeast; relatively low undercounts for post-strata with a relatively large Puerto Rican population; relatively high undercounts in some college towns; and overcounts in nonmetropolitan counties in the East North Central division. The majority does not conclude that any of these are sufficiently significant to undermine their overall face validity assessment.

As part of the accuracy discussion, the Committee also considered levels of sampling error. The evaluations had shown that the level of sampling error was higher than expected. Further work indicated that total variance might be underestimated because variance from the smoothing operation was not fully reflected. However, the majority concludes that the measured levels do not change their conclusions. There was also some discussion of the fact that most States are not different from the U.S. average at the 95-percent confidence level. However, the majority notes this might be expected in a tight distribution measured by a sample survey. Most States have a measure of undercount significantly different from zero. References to the total error model show that, for relative comparisons and especially for count comparisons, the adjusted figures are closer to the "truth" for more States. While analysis was not available for smaller areas, the majority concludes that acceptable patterns would happen there also.

Total Error Model and Loss Function Analysis. The results of the measurement of individual sources of error were entered into a total error model. The objective behind the total error model is to produce both a measure of the bias in the PES estimates and an interval that will contain the "true" level of undercount. These data can be used to produce an estimate of undercount corrected for biases.

Loss function analysis is a method of comparing the census counts and the selected PES counts to this new estimate of undercount to determine which is closer. This analysis shows that for 39 of 50 states and most places of 100,000 or more, the adjusted population shares are closer to the "truth" than are the census population shares. The majority noted that the 39 states gaining accuracy through adjustment include those with disproportionate numbers of hard-to-enumerate populations. The 11 States whose loss function analysis suggests lower accuracy with adjusted population shares still would gain in absolute numbers and incur only small relative losses. The minority remains concerned about the 11 States with reduced accuracy under adjustment. Another variation of the loss function analysis is to compute congressional apportionment by comparing "truth" to the distributions of (1) the selected PES, and (2) the census. This analysis shows that adjusted numbers are closer to the truth for apportionment than are unadjusted census numbers.

The majority judges that the improvement in counts on the average for the Nation, States, and places over 100,000 population outweighs the risk that the accuracy of adjusted counts might be less for small areas.

A minority has concerns about overreliance on the total error model. They question whether the individual components of the total error model accurately measure the error, whether all potential sources of error are included in the model, and whether the model for calculating the loss functions is appropriate.

The Committee discussed the fact that the component of model error from the smoothing model is probably understated because, in general, estimates of sampling error for these kinds of models do not fully reflect between-model error. The majority concludes that a fuller accounting of variance would not change their overall assessment.

There was a concern raised about the adequacy of the loss function analysis. An expanded loss function analysis to address this concern is planned but not yet available. The majority expects that the new loss function analysis would show some additional areas for which census counts are closer to "truth" than adjusted counts, but not enough to change their overall conclusion. They generally base this judgment on their reading of the variances attached to the components of the estimated total bias. The minority is concerned because this expanded loss function is not available for review and analysis.

GUIDELINE 2

Guideline 2 states that "The 1990 Census may be adjusted if the adjusted counts are consistent and complete across all jurisdictional levels: national, state, local and census block. The resulting counts must be of sufficient quality and level of detail to be usable for Congressional reapportionment and legislative redistricting, and for all other purposes and at all levels for which census counts are published."

Summary of Guideline 2.--With respect to this guideline, the Committee concentrated on information related to the "synthetic" assumption, that is, that the probability of being missed by the census is constant for each person within a particular post-stratum or that there should be homogeneity for block parts within the same post-stratum grouping. For analysis, the Committee relies on study P12 (evaluation of the synthetic assumption). The total error/loss function results were analyzed to support the P12 findings and to address the usability of adjusted data. Study results directly related to synthetic estimation were relatively limited. That was expected, since it is very difficult to get direct evidence for small areas. The majority of the Committee is sufficiently convinced from the data, that, in general, block parts are homogeneous within post-strata. They believe that despite some difficulties, adjusted counts (even for small areas) are better on average than census counts. They also see no evidence to indicate there are any serious flaws in the state synthetic assumption. The minority is concerned about the inconsistent findings in the P12 study. They are also concerned about the limited information available on the effect of the synthetic assumption on blocks.

Discussion.--The Committee notes that there is no PES system (except a second fully accurate census) that could say adjusted counts are more accurate for all blocks. The Committee recognizes that some blocks would be made worse by an adjustment. But the Committee believes this is acceptable if adjusted counts are on average more accurate than census counts for places and states. The Committee concludes that the accuracy of adjusted counts would improve for aggregations of blocks.

The synthetic estimation evaluation study (P12) presented information about whether or not block parts within the same post-strata were homogeneous. There were two types of analyses. One examined the homogeneity of state parts within post-strata. The other examined the homogeneity of blocks within post-strata. Both sets of analyses were subject to various limitations, and interpretation of the results would vary depending on the weight given to the limitations and the value placed on the methods.

The part of P12 that analyzed state homogeneity had two studies. One study used Analysis of Variance (ANOVA) to determine the validity of using post-strata, rather than state, for estimation. The study was designed to determine if there was more homogeneity within state or post-strata. By estimating if state differences were significant within post-strata, the study showed that, with the exception of the Mid Atlantic Division, they were not. This result supported the use of post-strata and convinces the majority that there is relative homogeneity for state parts within post-strata. They believe this result would hold for other geographic levels.

The second state homogeneity study examined six characteristics correlated with coverage error. Various tests for homogeneity were done that show that there is significant state effect within post-strata. While this result is undesirable, for the majority, the limitations were not evaluated as refuting the increased accuracy of the adjusted figures overall. For the majority of the Committee, there was more weight and value placed on the ANOVA state homogeneity study and the block homogeneity study (described below). The minority is concerned about the contradictory evidence of the two state studies.

The second part of P12 examined block homogeneity. Using a regression prediction model, adjustment factors were estimated at the block level, based on the distribution of the characteristics within individual blocks. The estimated or predicted factors were compared to the calculated factors assigned by the PES selected method. In a large majority of cases, the results validate the accuracy of the PES selected method.

Though not part of P12, the Committee also considered some theoretical work that indicated if biases and variances are low, and if there are reasonably behaved populations, adjusting at small levels will, on average, yield an improvement.

From these results, the majority concludes that block level coverage rates are homogeneous among blocks within post-strata, and, therefore, it is likely that characteristics related to coverage are also homogeneous. The minority is concerned about the unknown effect of the regression prediction model used in P12 and the fact that the study results are not consistent by census division. They also are concerned about limited evidence to indicate if the assumptions of the theoretical findings were met.

GUIDELINE 3

Guideline 3 states that "The 1990 census may be adjusted if the estimates generated from the pre-specified procedures that will lead to an adjustment decision are shown to be more accurate than the census enumeration. In particular, these estimates must be shown to be robust to variations in reasonable alternatives to the production figures, and to variations in the statistical models used to generate adjusted procedures." This guideline addresses two issues--having prespecified procedures and robustness.

Summary of Guideline 3.--With respect to prespecification, the Committee concludes that almost all aspects of the PES system were carried out as prespecified and those few that weren't precisely specified in advance were handled properly based on the best professional judgment. It would have been a mistake not to have corrected a major problem simply because the solution had not been fully prespecified.

With regard to robustness, the Committee is satisfied with the robustness of the imputation model. The Committee is sufficiently satisfied with the robustness of post-stratification variables, although the minority believe this is a concern. The Committee is concerned about the robustness of the smoothing model. For the majority, with reliance on the total error model and the analysis of confidence intervals by age, race, sex, and division, the concerns are not serious enough to change their overall conclusion that adjusted counts are more accurate. The minority disagrees on this point.

Prespecification.--The post-enumeration survey/adjustment process was carried out almost entirely as specified. The Committee identified only two possible deviations--the treatment of outliers in two sample segments and outliers in the variance smoothing model. With regard to outliers in the sample observations, it was prespecified that we would screen for and modify outlier points. There was no prespecified procedure for exactly how to modify the outlier points since there is no uniformly accepted practice within the statistical community. Two outlier sample segments were recognized and modified. All Committee members are satisfied that these outliers were handled in a reasonable manner and that the guideline of prespecification is met with regard to sample observation outliers.

Outliers in the variance smoothing model were not anticipated; therefore, there was no prespecified method for identifying them or for modifying them. This is the only instance where the Committee thinks that the PES/adjustment methodology did not follow prespecified procedures, but the Committee considers that this deviation was highly desirable. A description of this issue follows:

During the initial smoothing of the variance-covariance matrix of the raw factors, it became obvious that there were some unusual points. These were points with extraordinarily high undercount or overcount rates and high variances that, when the variances were lowered by smoothing would have been given more weight and would have distorted the adjustment numbers. Even though the data had already been seen, some rules were set up to identify outliers to be left out of the variance smoothing. In general, about four points per region were identified. After rerunning the smoothing procedures, it appeared that there were other outlier points. These were also removed, and the smoothing was done again. The smoothing with no outlier identification is called "modified" PES. The smoothing after removing the total set of outliers is called the "selected" PES, since the Committee preferred that method and selected it for producing adjusted counts. There is no name generally accepted for the smoothing that only removed the first four outliers. After the variance-covariance matrix was smoothed, the factors were smoothed. Outliers (though not very many) were also identified and adjusted in that smoothing process.

Even though variance outliers were not anticipated, the Committee judged it would be a mistake to ignore their effect, and therefore, selected a method for handling outliers even though the method was not prespecified. This approach was discussed with members of the Special Advisory Panel, who also agreed that outliers should be handled separately. The Committee, based on recommendations of the staff, strongly believes that the full outlier method (Selected PES) is preferable to the intermediate step (only four outliers), even though the two methods lead to differences that concern some members. (See Robustness)

Robustness.--With regard to robustness, the Committee investigated three operations: (1) the investigation of reasonable alternative imputation methods for missing data, (2) the post-stratification groupings, and (3) the smoothing model.

(1) Alternative Imputation Methods.--An imputation model was developed and prespecified to handle missing data in the PES. In addition, a simulation study was conducted to test the sensitivity of the effect of this model on the eventual dual system estimator (DSE). It shows a very low range of differences in the DSE due to the effect of missing data. The Committee unanimously agrees that the PES system is robust with regard to the imputation model for missing data, largely due to the low level of missing data.

(2) Post-stratification.--Once the PES data are matched to the census, DSE's are calculated within several subgroupings of the population called post-strata. A certain set of post-strata were prespecified (see appendix 5.) Census division was one of the items used to define post-strata. The assumption was that after accounting for race, age, sex, tenure, and place size, the undercount rate would be relatively homogeneous across states within a division. The raw PES factors showed that states within the same census division had very similar estimated undercount rates. To find out if the PES system of defining post-strata was robust, several analyses were done.

Even though the PES sample was not controlled by state, direct PES estimates for states were produced and compared to the preliminary PES estimates. Only 3 states show differences between the direct State and PES estimates at the 95 percent confidence interval, but they are all in one region, are contiguous, and show differences with the preliminary poststratified estimates all in the same direction. Other analyses were requested by the Committee during these deliberations, but were not completed in time for the Committee's decision.

From the data available, and making judgments about what could be shown in the yet to be completed analyses, the majority concludes that the post-strata grouping by division is not a major concern. The majority recognizes that different post-strata groupings would lead to different state estimates, but they judge the range would be small and that the different state estimates would still be more accurate than the census. This assessment relies on the analysis from a part of study P12 which shows that, except for one census division, post-strata have a stronger effect on undercount than individual states. From this, the majority concludes that any lack of robustness from alternative post-stratum groupings would not seriously affect the accuracy of adjusted counts. The minority notes that part of the P12 study shows a significant state effect. They also are concerned about the lack of evidence about whether alternative post-stratum groupings would be more accurate than census counts.

(3) Smoothing.--Once the dual system estimates were produced for each of the 1,392 post-strata, they were used to produce the preliminary or "raw" PES estimates. There was general prior agreement that the unsmoothed estimates would have large sampling errors and that some of these points would be unstable. Thus, there was a prespecified procedure to reduce the variance of the adjustment factors. That was given the name "smoothing." The first step was to smooth the variances (since the directly estimated variances were themselves estimates and subject to high instability). The next step was to use the smoothed variances and raw correlations to construct the presumed covariances of the raw factors. In turn, these covariances were incorporated into the multivariate regression model. The final step was to take a weighted average of the observed adjustment factor and the regression predicted adjustment factor. The results were controlled at the regional level. As mentioned, the smoothing model was prespecified but some of the variables were not. A set of ten were "must" variables, but others came in depending on the data. The method for choosing the other variables was prespecified.

When considering the outlier variance problem, the Committee noticed major differences in estimated smoothed variances between eliminating four outliers and eliminating seven outliers. The extent of differences from making a relatively minor change in assumptions is disturbing. A separate model was run to help understand the robustness of the prespecified procedure. It used as input 116 points which were a summary of the 1,392 points collapsed across age and sex (116=1392/12). It produced some different results, especially for the South Central and South Atlantic divisions. Because the results appeared sensitive to the variance smoothing, an alternative approach was attempted that smoothed not only the variances but the correlations; this procedure was run in the South and West and produced relatively large differences in the estimates. Another approach was also tried, based on summarizing the 1392 points into 116 points. Models for minorities and nonminorities produced estimates that broadly agreed with the results of the selected smoothing but differed by relatively large amounts relative to the uncertainty attributed to the selected model. There is also concern about the set of predictor variables, specifically that a different approach to selecting predictor variables might produce a noticeably different set of results. Analysis on this issue is also not available. However, the Committee notes that this sensitivity analysis should be expected to produce somewhat different results since it will be done after looking at the data and developing hypotheses about possible problems with the prespecified model.

Even though there are concerns about this apparent lack of robustness in the strictest sense, the majority of the Committee concludes that those concerns are not serious enough to change their overall conclusion. They based that conclusion on several factors:

1. All of the models produce dual system estimates (DSE's) that stand apart from the census counts in the same general direction. Members reference a June 6, 1991 special tabulation by Bell titled "Raw and Smoothed (from various models) Adjustment Factors and Census Adjustments."

2. Model error (both within the selected model and between potential models) is not fully reflected in the total error model (discussed under Guideline 1 above). However, sensitivity analyses conducted thus far do not suggest fatal flaws in the selected model adjustment.

3. All of the models, though producing somewhat different results, are doing what was intended--lessening the impact of unusual points by borrowing strength from other information.

In summary, the Committee is concerned about the lack of robustness in the strictest sense and potential problems in the smoothing process. On balance, the majority finds there is no evidence to conclude that concerns about the smoothing model would affect their overall assessment about the accuracy of adjusted numbers. Many believe that while it is likely a better model will be developed later, even one that would show adjustment to be less accurate for more States than the current results, that would not affect the judgment that adjusted counts by the selected model are more accurate than census counts.

The minority cannot conclude that lack of robustness in the smoothing model is a small enough problem not to affect the accuracy of the adjusted numbers. They base their opinion on: (1) A serious concern about the unknown effect of using unsmoothed covariances, (2) no evidence that the total error model includes the appropriate level of model error from the smoothing operation, (3) no evidence to be able to conclude that eventual information about other aspects of the model (for example: new predictor variables) would show the system to be robust, and (4) the substantial change in estimates when only a few observations are treated differently in smoothing.

GUIDELINE 6

Guideline 6 states that "There will be a determination whether to adjust the 1990 census when sufficient data are available, and when analysis of the data is complete enough to make such a determination. If sufficient data and analysis of the data are not available in time to publish adjusted counts by July 15, 1991, a determination will be made not to adjust the 1990 census."

As stated earlier in the introduction, the Committee decision is based on the data available and the methods chosen at this time. A large majority of members is comfortable that they have sufficient information to conclude that adjusted counts are more accurate than census counts. All members of the majority wish there had been more time to analyze more fully the data available, but they conclude in their best professional judgment and based on their examination of the study results that further analysis and future work will not lead to such different adjusted numbers that it would change their conclusion.

The work to date has raised questions that the Census Bureau is addressing in ongoing research. The Census Bureau has an obligation to continue its research into coverage measurement and adjustment procedures and will do so. Throughout the discussion of Guidelines 1-3, we have mentioned ongoing work that may lead to better understanding of some issues. We will not list those projects again here. Almost all of these ongoing projects are the result of questions and issues raised during the Committee's deliberations. As stated earlier, all of the prespecified 33 evaluations of the PES and demographic analysis were completed; as were a number of additional analyses that came up during the Committee's deliberations.

The minority believes it cannot conclude that adjusted counts are more accurate than census counts until it has had time to fully analyze the evaluation studies and their effect on the estimates.

APPENDIX 1: MEMBERS OF THE UNDERCOUNT STEERING COMMITTEE

Paula J. Schneider, Chair
Chief, Population Division

Peter A. Bounpane
Assistant Director, Decennial Census

William P. Butz
Associate Director for Demographic Programs

Robert Fay, Ph.D.
Senior Mathematical Statistician

Robert M. Groves, Ph.D.
Associate Director for Statistical Design, Methodology, and Standards

Charles D. Jones
Associate Director for Decennial Census

Susan M. Miskura
Chief, Year 2000 Research and Development Staff
Former Chief, Decennial Planning Division

John Thompson
Chief, Statistical Support Division

Robert D. Tortora, Ph.D.
Chief, Statistical Research Division

APPENDIX 2.

1. Table 1: Historical Estimates of the Amount and Percent of Net Undercount as Measured by Demographic Analysis, by Race: 1940 to 1990.

2. Table 2: Comparison of the Census, Post Enumeration Survey (PES) and Demographic Analysis (DA) Estimates of the Population and Percent Net Undercount: 1990

3. Table 3: Selected Post-Enumeration Survey (PES) Estimates of Total Resident Population: United States Total

4. Table 4: PES Estimates of Under/overcount Rates by State

Table 1. Historical Estimates of the Amount and Percent of Net Undercount as Measured by Demographic Analysis, by Race: 1940 to 1990

	Demographic Analysis Estimates of Net Undercount[1] (Amount in thousands)											
	1990		1980		1970		1960		1950		1940	
	Amount	%	Amount	%	Amount	%	Amount	%	Amount	%	Amount	%
Total Population	4,684	1.8	2,802	1.2	5,653	2.7	5,700	3.1	6,537	4.1	7,513	5.4
Black	1,836	5.7	1,257	4.5	1,566	6.5	1,327	6.6	1,225	7.5	1,187	8.4
Nonblack	2,848	1.3	1,545	0.8	4,087	2.2	4,374	2.7	5,312	3.8	6,326	5.0
Difference	NA	4.4	NA	3.7	NA	4.3	NA	3.9	NA	3.8	NA	3.4

[1] Estimates represent "point" estimates of net undercount for each census; the estimates are subject to qualifications regarding their accuracy as quantified by "uncertainty" ranges (not shown). The estimates for 1940-1980 are based in part of the "reverse projection" of the population aged 65 and over in 1990 using estimates of population change, which adds another component of error in those coverage estimates. The estimates represent revisions of previously published coverage estimates for 1940-1980. The estimates (especially for 1990) represent "work-in-progress" as of May 1991 and are subject to further revision as new data and research become available.

Table 2

Comparison of the Census, Post Enumeration Survey (PES) and Demographic Analysis (DA) Estimates of Population and Percent Net Undercount: 1990

Race, Sex, Age	Census Counts as Tabulated (used for PES) (1)	Census Counts as Modified (used for DA) (2)	PES Estimated Population (3)	DA Estimated Population (4)	Difference Between PES and Census (Net Undercount): Amount (5)=(3)-(1)	Difference Between PES and Census (Net Undercount): Percent (6)=5/3	Difference Between DA and Census (Net Undercount): Amount (7)=(4)-(2)	Difference Between DA and Census (Net Undercount): Percent (8)=7/4	Difference Between PES and DA Net Undercount[1]: Amounts (9)=7-5	Difference Between PES and DA Net Undercount[1]: Percents (10)=8-6	Difference Between PES and DA Estimated Population[1]: Amount (11)=4-3	Difference Between PES and DA Estimated Population[1]: Percent (12)=11/4
TOTAL	248,709,873	248,709,873	253,979,140	253,393,786	5,269,267	2.07	4,683,913	1.85	(585,354)	-0.23	(585,354)	-0.23
Male	121,239,418	121,239,348	124,249,092	124,719,564	3,009,674	2.42	3,480,216	2.79	470,542	0.37	470,472	0.38
Female	127,470,455	127,470,525	129,730,048	128,674,222	2,259,593	1.74	1,203,697	0.94	(1,055,896)	-0.81	(1,055,826)	-0.82
BLACK	29,986,060	30,483,281	31,505,836	32,319,553	1,519,776	4.82	1,836,272	5.68	316,496	0.86	813,717	2.52
Male	14,170,151	14,420,331	14,974,384	15,758,711	804,233	5.37	1,338,380	8.49	534,147	3.12	784,327	4.98
Female	15,815,909	16,062,950	16,531,452	16,560,842	715,543	4.33	497,892	3.01	(217,651)	-1.32	29,390	0.18
NONBLACK	218,723,813	218,226,592	222,473,304	221,074,233	3,749,491	1.69	2,847,641	1.29	(901,850)	-0.40	(1,399,071)	-0.63
Male	107,069,267	106,819,017	109,274,708	108,960,853	2,205,441	2.02	2,141,836	1.97	(63,605)	-0.05	(313,855)	-0.29
Female	111,654,546	111,407,575	113,198,596	112,113,380	1,544,050	1.36	705,805	0.63	(838,245)	-0.73	(1,085,216)	-0.97
TOTAL MALE												
All ages	121,239,418	121,239,348	124,249,092	124,719,564	3,009,674	2.42	3,480,216	2.79	470,542	0.37	470,472	0.38
0-9	18,654,936	18,831,059	19,432,751	19,525,522	777,815	4.00	694,463	3.56	(83,352)	-0.45	92,771	0.48
10-19	17,869,865	17,911,634	18,162,241	17,829,952	292,376	1.61	(81,682)	-0.46	(374,058)	-2.07	(332,289)	-1.86
20-29	20,371,532	20,445,048	21,472,875	21,093,431	1,101,343	5.13	648,383	3.07	(452,960)	-2.06	(379,444)	-1.80
30-44	29,471,160	29,371,471	30,259,834	30,553,084	788,674	2.61	1,181,613	3.87	392,939	1.26	293,250	0.96
45-64	22,306,752	22,187,370	22,447,849	23,031,859	141,097	0.63	844,489	3.67	703,392	3.04	584,010	2.54
65+	12,565,173	12,492,766	12,473,542	12,685,716	(91,631)	-0.73	192,950	1.52	284,581	2.26	212,174	1.67
TOTAL FEMALE												
All ages	127,470,455	127,470,525	129,730,048	128,674,222	2,259,593	1.74	1,203,697	0.94	(1,055,896)	-0.81	(1,055,826)	-0.82
0-9	17,798,686	17,961,366	18,553,839	18,640,971	755,153	4.07	679,605	3.65	(75,548)	-0.42	87,132	0.47
10-19	16,998,399	17,030,546	17,373,452	17,023,405	375,053	2.16	(7,141)	-0.04	(382,194)	-2.20	(350,047)	-2.06
20-29	19,961,825	20,014,399	20,829,943	20,272,650	868,118	4.17	258,251	1.27	(609,867)	-2.89	(557,293)	-2.75
30-44	29,970,630	29,896,153	30,477,082	30,022,880	506,452	1.66	126,727	0.42	(379,725)	-1.24	(454,202)	-1.51
45-64	24,064,257	23,981,932	24,012,341	24,079,786	(51,916)	-0.22	97,854	0.41	149,770	0.62	67,445	0.28
65+	18,676,658	18,586,129	18,483,391	18,634,530	(193,267)	-1.05	48,401	0.26	241,668	1.31	151,139	0.81

[1] A positive difference means that the demographic estimate is higher than the PES estimate; a negative difference means that the demographic estimate is lower.

Table 2 (Cont'd)

Comparison of the Census, Post Enumeration Survey (PES) and Demographic Analysis (DA) Estimates of Population and Percent Net Undercount: 1990

Race, Sex, Age	Census Counts as Tabulated (used for PES) (1)	Census Counts as Modified (used for DA) (2)	PES Estimated Population (3)	DA Estimated Population (4)	PES Net Undercount: Amount (5)=(3)-(1)	PES Net Undercount: Percent (6)=5/3	DA Net Undercount: Amount (7)=(4)-(2)	DA Net Undercount: Percent (8)=7/4	Difference Between PES and DA Net Undercount[1/]: Amounts (9)=7-5	Difference Between PES and DA Net Undercount[1/]: Percents (10)=8-6	Difference Between PES and DA Estimated Population[1/]: Amount (11)=4-3	Difference Between PES and DA Estimated Population[1/]: Percent (12)=11/4
BLACK MALE												
All ages	14,170,151	14,420,331	14,974,384	15,758,711	804,233	5.37	1,338,380	8.49	534,147	3.12	784,327	4.98
0-9	2,758,760	2,857,825	2,998,847	3,112,136	240,087	8.01	254,311	8.17	14,224	0.17	113,289	3.64
10-19	2,656,671	2,698,565	2,768,729	2,753,791	112,058	4.05	55,226	2.01	(56,832)	-2.04	(14,938)	-0.54
20-29	2,544,386	2,621,647	2,718,157	2,891,989	173,771	6.39	270,342	9.35	96,571	2.95	173,832	6.01
30-44	3,199,321	3,232,061	3,399,720	3,689,721	200,399	5.89	457,660	12.40	257,261	6.51	290,001	7.86
45-64	2,045,581	2,053,297	2,113,856	2,324,165	68,275	3.23	270,868	11.65	202,593	8.42	210,309	9.05
65+	965,432	956,936	975,075	986,909	9,643	0.99	29,973	3.04	20,330	2.05	11,834	1.20
BLACK FEMALE												
All ages	15,815,909	16,062,950	16,531,452	16,560,842	715,543	4.33	497,892	3.01	(217,651)	-1.32	29,390	0.18
0-9	2,698,251	2,792,907	2,926,950	3,030,650	228,699	7.81	237,743	7.84	9,044	0.03	103,700	3.42
10-19	2,603,412	2,645,152	2,712,116	2,705,551	108,704	4.01	60,399	2.23	(48,305)	-1.78	(6,565)	-0.24
20-29	2,742,332	2,812,858	2,942,171	2,922,200	199,839	6.79	109,342	3.74	(90,497)	-3.05	(19,971)	-0.68
30-44	3,695,231	3,726,605	3,845,110	3,822,938	149,879	3.90	96,333	2.52	(53,546)	-1.38	(22,172)	-0.58
45-64	2,533,564	2,550,143	2,566,637	2,563,789	33,073	1.29	13,646	0.53	(19,427)	-0.76	(2,848)	-0.11
65+	1,543,119	1,535,285	1,538,468	1,515,714	(4,651)	-0.30	(19,571)	-1.29	(14,920)	-0.99	(22,754)	-1.50
NONBLACK MALE												
All ages	107,069,267	106,819,017	109,274,708	108,960,853	2,205,441	2.02	2,141,836	1.97	(63,605)	-0.05	(313,855)	-0.29
0-9	15,896,178	15,973,234	16,433,904	16,413,386	537,728	3.27	440,152	2.68	(97,576)	-0.59	(20,518)	-0.13
10-19	15,213,194	15,213,069	15,393,512	15,076,161	180,318	1.17	(136,908)	-0.91	(317,226)	-2.08	(317,351)	-2.10
20-29	17,827,146	17,823,401	18,754,718	18,201,442	927,572	4.95	378,041	2.08	(549,531)	-2.87	(553,276)	-3.04
30-44	26,271,839	26,139,410	26,860,114	26,863,363	588,275	2.19	723,953	2.69	135,678	0.50	3,249	0.01
45-64	20,261,171	20,134,073	20,333,993	20,707,694	72,822	0.36	573,621	2.77	500,799	2.41	373,701	1.80
65+	11,599,741	11,535,830	11,498,467	11,698,807	(101,274)	-0.88	162,977	1.39	264,251	2.27	200,340	1.71
NONBLACK FEMALE												
All ages	111,654,546	111,407,575	113,198,596	112,113,380	1,544,050	1.36	705,805	0.63	(838,245)	-0.73	(1,085,216)	-0.97
0-9	15,100,435	15,168,459	15,626,889	15,610,321	526,454	3.37	441,862	2.83	(84,592)	-0.54	(16,568)	-0.11
10-19	14,394,987	14,385,394	14,661,336	14,317,854	266,349	1.82	(67,540)	-0.47	(333,889)	-2.29	(343,482)	-2.40
20-29	17,219,493	17,201,541	17,887,772	17,350,450	668,279	3.74	148,909	0.86	(519,370)	-2.88	(537,322)	-3.10
30-44	26,275,399	26,169,548	26,631,972	26,199,942	356,573	1.36	30,394	0.12	(326,179)	-1.22	(432,030)	-1.65
45-64	21,530,693	21,431,789	21,443,784	21,515,997	(86,909)	-0.40	84,208	0.39	169,197	0.79	72,273	0.33
65+	17,133,539	17,050,844	16,944,923	17,118,816	(188,616)	-1.11	67,972	0.40	256,588	1.51	173,893	1.02

[1/] A positive difference means that the demographic estimate is higher than the PES estimate; a negative difference means that the demographic estimate is lower.

Table 3. Selected Post-Enumeration Survey (PES) Estimates of Total Resident Population: United States Total

Race/Hispanic/ Sex Group	Resident Census Enumeration[1] (1)	Selected PES Estimate of Population (Rounded) (2)	Estimated Under/Over Count Rate[2] (3)	Margin of Error due to Sampling[3] (4)
Total	248,709,873	253,979,000	2.1	0.4
Male	121,239,418	124,249,000	2.4	0.4
Female	127,470,455	129,730,000	1.7	0.4
Black	29,986,060	31,505,000	4.8	0.6
Male	14,170,151	14,974,000	5.4	0.6
Female	15,815,909	16,531,000	4.3	0.6
Non-Black	218,723,813	222,474,000	1.7	0.4
Male	107,069,267	109,275,000	2.0	0.4
Female	111,654,546	113,199,000	1.4	0.4
Other Populations of Interest				
Asian or Pacific Islander	7,273,662	7,504,000	3.1	0.9
Male	3,558,038	3,688,000	3.5	1.0
Female	3,715,624	3,816,000	2.6	0.9
American Indian	1,878,285	1,977,000	5.0	2.1
Male	926,056	981,000	5.6	2.2
Female	952,229	996,000	4.4	2.0
Hispanic[4]	22,354,059	23,591,000	5.2	0.8
Male	11,388,059	12,087,000	5.8	0.9
Female	10,966,000	11,504,000	4.7	0.9

Note: Estimate of total population may differ from other tables due to rounding.

[1] The population counts released are subject to possible correction for undercount or overcount. The United States Department of Commerce is considering whether to correct the counts and will publish corrected counts, if any, no later than July 15, 1991.

[2] Negative values indicate an overcount.

[3] Add to and subtract from estimated under/over count rate to obtain a 95% confidence interval.

[4] Persons of Hispanic Origin may be any race.

Table 4. PES Estimates of Under/Over Count Rates by State

State	Resident Census Enumeration[1]	Selected PES Method		
		Estimated Population (Rounded)	Under/ Over Count Rate	Margin of Error[2]
United States Total	248,709,873	253,978,000	2.1	0.4
Alabama	4,040,587	4,146,000	2.5	0.8
Alaska	550,043	561,000	2.0	0.9
Arizona	3,665,228	3,790,000	3.3	0.9
Arkansas	2,350,725	2,403,000	2.2	0.8
California	29,760,021	30,888,000	3.7	0.8
Colorado	3,294,394	3,376,000	2.4	0.9
Connecticut	3,287,116	3,306,000	0.6	1.1
Delaware	666,168	687,000	3.0	0.9
District of Columbia	606,900	639,000	5.0	1.0
Florida	12,937,926	13,278,000	2.6	0.8
Georgia	6,478,216	6,633,000	2.3	0.7
Hawaii	1,108,229	1,136,000	2.4	1.1
Idaho	1,006,749	1,035,000	2.7	1.0
Illinois	11,430,602	11,592,000	1.4	0.7
Indiana	5,544,159	5,586,000	0.7	0.7
Iowa	2,776,755	2,807,000	1.1	0.9
Kansas	2,477,574	2,506,000	1.1	0.7
Kentucky	3,685,296	3,768,000	2.2	0.8
Louisiana	4,219,973	4,332,000	2.6	0.7
Maine	1,227,928	1,240,000	1.0	1.2
Maryland	4,781,468	4,869,000	1.8	0.9
Massachusetts	6,016,425	6,039,000	0.4	1.1
Michigan	9,295,297	9,404,000	1.2	0.7
Minnesota	4,375,099	4,419,000	1.0	0.7
Mississippi	2,573,216	2,632,000	2.2	0.8
Missouri	5,117,073	5,184,000	1.3	0.7
Montana	799,065	822,000	2.8	1.0
Nebraska	1,578,385	1,595,000	1.0	0.8
Nevada	1,201,833	1,232,000	2.4	0.9
New Hampshire	1,109,252	1,116,000	0.6	1.1
New Jersey	7,730,188	7,836,000	1.4	1.0
New Mexico	1,515,069	1,586,000	4.5	1.0
New York	17,990,455	18,304,000	1.7	0.9
North Carolina	6,628,637	6,815,000	2.7	0.7
North Dakota	638,800	648,000	1.4	0.9
Ohio	10,847,115	10,933,000	0.8	0.7
Oklahoma	3,145,585	3,214,000	2.1	0.8
Oregon	2,842,321	2,898,000	1.9	0.9
Pennsylvania	11,881,643	11,957,000	0.6	1.0
Rhode Island	1,003,464	1,006,000	0.3	1.1
South Carolina	3,486,703	3,590,000	2.9	0.8
South Dakota	696,004	707,000	1.6	1.0
Tennessee	4,877,185	5,012,000	2.7	0.8
Texas	16,986,510	17,551,000	3.2	0.8
Utah	1,722,850	1,757,000	1.9	1.1
Vermont	562,758	571,000	1.4	1.4
Virginia	6,187,358	6,353,000	2.6	0.7
Washington	4,866,692	4,987,000	2.4	0.9
West Virginia	1,793,477	1,842,000	2.6	0.9
Wisconsin	4,891,769	4,924,000	0.7	0.7
Wyoming	453,588	466,000	2.7	1.0

[1] The population counts released are subject to possible correction for undercount or overcount. The United States Department of Commerce is considering whether to correct the counts and will publish corrected counts, if any, no later than July 15, 1991.

[2] Add to and subtract from estimated under/over count rate to obtain a 95% confidence interval.

APPENDIX 3: BACKGROUND ON 1990 COVERAGE MEASUREMENT PROGRAMS AND ADJUSTMENT

Since the very first census, there have been problems in accurately counting every person living in the United States. The resulting undercount, or percentage of the population that is not counted by the census, is not a new phenomenon. Beginning with the 1940 census, each census has included an evaluation and research program to measure coverage error. The main purpose of these studies has been to measure progress in reducing the errors and to design programs to correct enumeration problems for the next census. In other words, the Census Bureau evaluated the most recent past census in order to improve the next census. In general, evaluation programs also provide information on limitations of the census being evaluated.

These evaluations showed a steady improvement in net census coverage over four decades, from an estimated undercount of more than 5 percent for the total population in 1940 to an estimated undercount in 1980 of just over 1 percent. They have also shown larger undercount rates for Blacks than nonBlacks and a differential that has stayed about 3-4 percentage points over the period.

Because of concern about this differential undercount, some people argued for the 1980 and 1990 censuses that if we can estimate the number of people missed in a census, why not simply correct the census to account for missed persons and make it more accurate. This, in simple terms, is what is called "adjustment." But estimating the census undercount with minimal error and, in turn, using that knowledge to improve the counts are two highly complex and difficult tasks.

The Census Bureau could not use the results of the 1980 census coverage evaluation programs to adjust the 1980 census counts. The Bureau considered the estimates of undercount to be flawed. It did not adjust the counts because staff were not confident that counts could be produced through the adjustment process that were better than the census counts. The coverage measurement methods in 1980 did not produce acceptable estimates of population or undercount on a national, state-by-state, or small area-basis.

For the 1990 census, the Bureau again conducted coverage measurement programs, built in improvements to address some of the problems experienced in 1980, and conducted 33 evaluation projects to determine how accurate the coverage measurement programs were.

The Census Bureau had two major programs to measure coverage in the 1990 census. The first was the post-enumeration survey (PES), which was a sample survey we took after the census. Approximately 165,000 housing units in a sample of 5,290 census blocks or block clusters were interviewed. Persons enumerated during the PES were also referred to as the P-sample. After persons in the housing units were interviewed, their forms were matched to census records to determine whether they were counted in the census. This process measured gross <u>erroneous omissions</u> in the census.

The Bureau also measured gross <u>erroneous inclusions</u> in the census by determining whether any of the persons in the sample blocks who were enumerated in the census should not have been counted there. An erroneous census enumeration, for example, could have included a child born after April 1, 1990, a person who died before April 1, or a college student away from home who was enumerated at his or her parents' address, instead of being correctly enumerated at the college. Persons who were enumerated in or otherwise coded to the PES sample blocks, even if coded in error, constitute the E-sample.

The data on gross erroneous inclusions and gross erroneous omissions were used to produce an estimate of the net undercount or net overcount of the population in the census. This was a very complex process that combined elements of survey design, interviewing, matching, imputation, mathematical modeling and professional judgment.

Second, the Census Bureau used demographic analysis estimates, which were produced by combining various sources of administrative data, to produce an estimate of total population. This included using historical data on births, deaths, and legal immigration; estimates of emigration and undocumented immigration; and Medicare data.

Demographic analysis estimates were used to evaluate the reasonableness of the PES estimates. Only the PES provided estimates of undercount and overcount at a level of detail suitable for use in adjustment. For example, demographic analysis estimates were produced only at the national level and for Black and nonBlack; the PES process was designed to measure coverage error for more population groups (Whites, Blacks, Hispanics, Asians and Pacific Islanders, and American Indians) by detailed levels of geography. Therefore, only the PES data would permit an adjustment.

We will discuss each of these programs now. For a more detailed discussion of PES see Howard Hogan, "The 1990 Post-Enumeration Survey: An Overview," a paper presented at the American Statistical Association in August 1990; for a more detailed discussion of demographic analysis see J. Gregory Robinson, "Plans for Estimating Coverage of the 1990 United States Census: Demographic Analysis," a paper presented to the Southern Demographic Association, in October, 1989.

POST-ENUMERATION SURVEY (PES)

Listing and Interviewing

Post-enumeration survey sample blocks were chosen to provide enough information to estimate the undercount or overcount for some 1,392 distinct groups of the population. (See appendix 5.) These groups are called post-strata and will be described later, but every person resident in the United States belongs in one of the post-strata except for those persons who were out-of-scope in the PES. Persons living in institutions were excluded from the PES, as were military personnel living in barracks, people living in remote rural Alaska, and persons in emergency shelters and persons who had no formal shelter.

In February 1990, Census Bureau current survey enumerators visited each of the sample blocks to list all housing units they contained. To preserve independence, none of the temporary enumerators hired to take the 1990 census was used nor was the listing conducted out of the temporary census offices. To maintain independence, the Census Bureau did not want anyone to know where a PES sample block was so that it would be treated differently during the census.

After the completion of the 1990 census follow up of those housing units that did not return a form (called nonresponse followup), enumerators interviewed persons at households in the PES sample blocks. Although this interviewing drew from enumerators who had worked on 1990 census follow up, steps were taken to preserve independence, such as not allowing an enumerator to work in a block in the PES that he or she had worked in during the census.

The interviewers determined who was living in each housing unit, obtained their characteristics, and asked where they lived on April 1, 1990, Census Day. The PES interviewing began nearly 3 months after Census Day. Many people had moved during that time. In order to determine whether they were enumerated in the census, the Bureau needed to know where they lived on Census Day and, thus, enumerators asked a series of probing questions to determine occupants' Census Day addresses.

There was a quality assurance program for the interviewing phase to ensure that the interviewers really visited the household and that the people listed were indeed real. If interviewers made up people, they would not match to the census and would inflate the undercount rate.

Matching

The next step was to match the persons enumerated during the PES (the P-sample) to the census. The matching operation was the first step in determining whether persons in the P-sample (see page 2 for description) were enumerated by the census or missed. Ultimately, those persons in the P-sample matched to the census were considered to have been enumerated; those nonmatched were considered to have been missed.

Matching was carried out in four stages. It involved an initial stage of computer matching followed by two stages of clerical matching to attempt to resolve cases that the computer could not match. The two stages of clerical matching were differentiated by the level of skill and judgment required to establish a match.

Those persons in the P-sample not matched to the census by computer and the first two stages of clerical matching were assigned for a followup interview, if it was determined that additional information was necessary to establish whether a match to the census was appropriate. An additional fourth stage of clerical matching was then conducted that allowed the more skilled clerical matchers to use the information from the followup interview to establish additional matches.

The E-sample, those persons in the PES blocks who were enumerated in the census, was examined to determine if they were correctly enumerated. E-sample persons were matched back into the census to determine if they were enumerated more than once (duplicates). E-sample persons who were matched to the P-sample were assumed to be correctly enumerated (except for duplicate census enumerations). The remaining E-sample persons who were not matched to the P-sample were potential candidates for erroneous enumerations. These unmatched census persons were also included in the PES followup operation described above. The followup interviewers determined the enumeration status of those persons; that is, if they were correctly enumerated and simply not in the P-sample or if they were erroneously enumerated.

A final matching and reconciliation operation took place at the conclusion of the PES followup. This included the fourth stage of clerical matching for the P-sample and a determination of whether persons in the E-sample were correctly or erroneously enumerated. An important aspect of this operation was that situations arose where correct match status for persons in the P-sample, or correct enumeration status for persons in the E-sample, could not be determined. This situation occurred because the initial interview was inconclusive or because an incomplete interview was obtained during the followup.

Imputation and Dual System Estimates

A final PES file was created that reflected the results of the operations described above. This file included the characteristics of each person in the P-sample and the E-sample. The file also included the match status for persons in the P-sample and the enumeration status (correct or erroneous) for persons in the E-sample. As the final file was prepared, computer editing or imputation was performed to correct, insofar as possible, for missing or contradictory data. A critical aspect of imputation involved the estimation of a final match status for those persons whose match status could not otherwise be resolved. The estimation of match status was very critical. For example, mistakes in the PES matching process, which incorrectly identified persons as not counted in the census (nonmatches), erroneously overstated the estimated undercount.

The data in the final PES file were then summarized and incorporated with data from the full census to produce dual system (PES and census) estimates (DSE's) of total population. The DSE's were produced for unique estimation strata (or groupings of persons described below). The dual system estimator is explained more fully in Hogan's document cited above. Essentially it involves determining how many people were (1) in the PES and in the census, (2) in the PES and out of the census, (3) in the census but not in the PES, and (4) in neither the census nor PES.

Post-Strata

Using the match status and key data, such as age, race, and sex for each person in the sample, the Bureau began preparing the dual system estimates of the total population for each of 1,392 groupings of people. We call these groupings post-strata. The reason for forming the post-strata was to group persons who had similar chances of being enumerated in the census. The post-strata were defined by census division, geographic subdivisions such as central cities of large metropolitan statistical areas, whether the person was the owner or renter of the housing unit, race, age, and sex. Each person in the PES sample belonged in one of the unique post-strata. (For a list of post-strata see appendix 5.)

For purposes of illustration, the following are examples of the 1,392 post-strata. One example is a post-stratum which contains Black males, age 20-29, living in rented housing in central cities in the New York primary metropolitan statistical area. A second example is that which contains non-black non-Hispanic females, age 45-64, living in owned or rented housing in a non-metropolitan place of 10,000 or more population in the Mountain Division. A third example is that which contains Asian males, age 45-64, living in owned or rented housing in metropolitan statistical areas but not in a central city in the Pacific Division. A fourth example is that which contains non-black Hispanic females, age 30-44, living in owned or rented housing in central cities in the Los Angeles-Long Beach primary metropolitan statistical area or other central cities in metropolitan statistical areas in the Pacific region. As can be seen from these examples, the 1,392 post-strata are very specific.

Adjustment Factors

The next step in the post-enumeration survey process was to compare the estimated total population for each post-stratum (the dual system estimate or DSE) to the census count to determine a "raw" adjustment factor. For example, if the DSE for a particular post-stratum was 1,050,000 and the census count was 1,000,000, then the adjustment factor was 1.05, reflecting about a 5 percent estimated net undercount, with a range (plus or minus) of variability. An adjustment factor may be less than one, thus lowering the count in a post-stratum if an adjustment is applied. This results when there is evidence of an overcount in the post-stratum.

"Smoothing" the Adjustment Factors

The next step was "smoothing" these "raw" adjustment factors to reduce sampling variance associated with them and to produce final adjustment factors. Because the PES was a sample, it was subject to sampling error. Sampling error is an estimate of the error associated with taking some of the population (a sample) rather than all of the population (a census). The process of smoothing the "raw" adjustment factors to create final adjustment factors was a step to minimize the effect of sampling error. Of course, the smoothing technique was based on certain assumptions. The Bureau hoped that the gains from smoothing offset any potential errors in the smoothing model chosen. If the Bureau had not controlled by smoothing, the final adjustment factors for some of the post-strata would have been based on estimates of undercount that were subject to very large sampling error.

Small Area Estimation

The Census Bureau used the final adjustment factors to produce adjusted counts for every block in the Nation. The PES can only produce "direct" estimates of the total population for relatively large geographic areas (*i.e.*, the 1,392 post-strata). If there is a decision to adjust, however, the adjustment must be applied to each of the Nation's 4 million populated blocks. The Bureau developed a model that takes the adjustment factors produced for each of the 1,392 post-strata areas and uses them to estimate adjustment counts for each block. Since each of the post-strata crosses many blocks, the Bureau based its model on a critical assumption that coverage error is similar for all blocks that a post-stratum crosses.

Here are two examples of how block counts could be changed during this process. Suppose a census block with 200 people had 50 people who fell into a particular post-stratum. An adjustment factor of 1.05 was computed for that post-stratum, so 50 was multiplied 1.05, which comes to 52.5. Since procedures allowed adding only whole persons to a block, either 2 or 3 persons were added, based on a pre-specified procedure, to the persons in that post-stratum for that block. Other groupings of persons in the block in this example also were multiplied by the adjustment factor for the post-stratum into which they fell. Similarly, suppose there were 80 people in another post-stratum in a particular census block, and the adjustment factor was 0.94, indicating an overcount. 80 was multiplied by 0.94, which came to 75.2, so 4 or 5 person records were eliminated from that block.

The Bureau then produced a set of census tabulations with adjusted counts. It did this by adding or subtracting "adjustment" persons with detailed characteristics. The "adjusted" data files could then be used to produce all required census tabulations.

Evaluations

The Bureau conducted 22 evaluation studies that addressed various sources of potential error in the post-enumeration survey. The results of these evaluations were essential to determining whether adjusted or unadjusted census counts are more accurate.

In 1980, problems in the PES process resulted in data that were not suitable for adjusting the census. The Census Bureau's test census program for the 1990 census indicated that the effect of errors in the 1990 PES system could also be serious. The Bureau implemented new procedures for 1990 that made progress toward controlling errors in the PES. However, it conducted a careful evaluation of the PES to determine whether it had successfully controlled the errors that could occur in the PES process. The evaluations included assessments of a variety of problems including:

Missing Data

Missing data are imputed using statistical models to predict missing values. An important type of missing data occurs for those persons in the PES whose match status could not be resolved through our regular matching and interview efforts.

For the 1980 census, high rates of missing data in the PES contributed to uncertainty in coverage estimates. This uncertainty was one of the reasons for the Census Bureau's recommendation that 1980 coverage estimates were not sufficiently accurate to adjust that census.

The Bureau did everything it could to minimize the number of missing data cases in the 1990 PES. Special programs undertaken to reduce the level of missing data included numerous call-backs during the interviewing phase of the PES, as well as sending experienced interviewers back into the field in selected district office areas to attempt to resolve noninterview cases.

Matching

Matching of the census and the PES is also a critical part of the PES process. The Bureau must insure that matching has been accurately carried out since errors in matching have a direct effect on the accuracy of estimates of coverage error. Errors that lead to a failure to match persons in the P-sample to the census will lead to dual system estimates that overstate the correct population and overstate an estimated undercount.

The clerical stages of matching can be difficult and time-consuming and in certain instances must substantially rely on the judgment of the individual clerical matcher. For example, in the best case, both the census and the PES would record John Brown. But, are John Brown and J. Brown the same person? Another category of cases arises from different spellings. Yet another category of cases arises from similar names. Are Linda Martin and Linda Martinez the same person?

<u>Census Day Addresses</u>

Accurate reporting of census day addresses is a critical part of the PES. If accurate responses are not obtained on census day addresses, then persons would incorrectly be identified as missed in the census or incorrectly counted in the census would fail to be recognized.

<u>Persons Missed in the PES and the Census</u>

There are people who are extremely hard to count and are, consequently, likely to be missed in both the census and the PES. The PES process attempts to estimate these people. However, because there is no way to observe them directly, an indirect method was used to estimate their number. This method is based on assumptions. To the extent the assumptions underlying the estimation process are not fully satisfied, the PES estimates will be biased. Of even greater concern is that this type of error can occur differentially by geographic area or population group. Therefore, it is critical to evaluate the effect of this type of error.

DEMOGRAPHIC ANALYSIS

The Census Bureau's other coverage measurement program was demographic analysis (DA). The demographic coverage estimates could only be used to evaluate the completeness of coverage of the 1990 census at a national level and only for race (Black/Non-Black), sex, and age groups. DA could not provide even reasonably reliable coverage estimates for the Hispanic, Asian/Pacific Islander, or American Indian/Native Alaskan populations because these characteristics are often not recorded on birth and death certificates; nor can it provide estimates at the State or substate level. However, the PES measured under or overcounts of these groups. The DA coverage estimates were compared to the post-enumeration survey coverage estimates to assess the overall consistency of the two sets of estimates at the national level.

As mentioned above, DA used historical data on births, deaths, and legal immigration; estimates of emigration and undocumented immigration; and medicare data. Birth and death records are available for the entire United States from 1933 on, but are not complete for years before 1933. Therefore, the Bureau had to find other ways to estimate the number of people who were born or died prior to 1933. In estimating births for each year, The Bureau added to the number of registered births an estimate of underregistration. Underregistration was estimated based on tests conducted in 1940, 1950, and 1964-1968. If the estimates of underregistration are off, they could have a significant effect on undercount estimates because birth data are by far the largest component in estimating the population through demographic analysis.

The United States does not keep emigration records. Therefore, an estimate had to be made of those who have left the country. While the United States does have good records of legal immigration, there is no accurate estimate of illegal immigration. The Immigration and Naturalization Service now collects different information than it did prior to 1980. That change further complicated the effort to estimate legal immigration. Although recent legislative reform allowed nearly 3 million undocumented aliens to receive amnesty, the Bureau did not know whether all of these persons actually reside in the United States. One million could be migrant workers who actually live outside the country. Once again, The Bureau used professional judgment to estimate these components.

It is important to emphasize that results of demographic analysis are not exact but are estimates. To a large extent, they were based on assumptions and best professional judgment. As in the PES, the Bureau tried to estimate potential error in the data produced by demographic analysis. To estimate that overall error, the Bureau conducted 11 demographic analysis evaluation studies to find out as much as possible about each possible source of error. Based on these studies, the Bureau developed a range of error around the demographic analysis estimates.

APPENDIX 4. LIST OF POST-ENUMERATION SURVEY AND DEMOGRAPHIC ANALYSIS EVALUATION STUDIES

PES Studies

P1: Analysis of Reasonable Alternatives

P2: Distribution of Missing Data Rates

P3: Evaluation of Imputation Methodology for Unresolved Match Status Cases

P4: Quality of Reported Census Day Address--Evaluation Followup

P5: Analysis of PES Fabrications from Quality Control

P5A: Analysis of P-Sample Fabrication from Evaluation Followup Data

P6: Fabrication in the P-Sample--Interviewer Effects

P7: Estimates of Clerical Matching Error from the Evaluation

P8: Matching Error--Estimates of Clerical Matching Error in the P-Sample from Quality Assurance Results

P9: Accurate Measurement of Census Erroneous Enumerations--Consistency Checks

P9A: Accurate Measurement of Census Erroneous Enumerations--Evaluation Followup

P10: Accurate Measurement of Census Erroneous Enumerations--Clerical Error in Assignment of Census Enumeration Status

P11: Balancing Error Evaluation--Percentage of Matches Found Outside Sample Blocks

P12: Evaluation of the Synthetic Assumption

P13: Evidence of Correlation Bias from Demographic Analysis

P13, Part 2: Evidence of Correlation Bias--Alho Technique

P14: Independence of the Census and the P-sample

P14, Part 2: Debriefing of PES's Interviewers

P15: Random Error of Coverage Estimates

P16: Total Error

P17: Internal Consistency of Estimates

P18: Evaluation of Late/Late Census Data

DA Studies

D1: Error in Birth Underregistration Completeness Estimates

D2: Uncertainty in Estimates of Undocumented Aliens

D3: Uncertainty in Estimated White Births, 1915-1935

D4: Uncertainty in Estimated Black Births, 1915-1935

D5: Robustness of Estimated Number of Emigrants

D6: Robustness of Estimates of the Population 65 and Older

D7: Uncertainty Measures for Other Components

D8: Uncertainty of Models to Translate 1990 Census Concepts into Historical Racial Classifications

D9: Inconsistencies in Race Classifications of the Demographic Estimates and the Census

D10: Differences Between Preliminary and Final Demographic Estimates

D11: Total Error in the Demographic Estimates

APPENDIX 5. POST-ENUMERATION SURVEY POST-STRATA

The 1990 Post-Enumeration Survey (PES) provides direct estimates for 1,392 post-strata. The post-strata are designed to divide the PES sample blocks into groups which have similar characteristics. This helps the Census Bureau to estimate the coverage of the 1990 decennial census more accurately.

The post-strata are defined by census division, area (city, non-city, rural, etc.), race, Hispanic origin, tenure group, sex, and age. Tenure refers to whether housing units are owned or rented. Each post-strata is given an eight digit code. The attached document shows 116 post-strata grouped by evaluation post-strata and the corresponding first six digits of the post-stratum code for each. The last two digits are not delineated on the attachment. They define sex and age group. There are six age group classifications and two sex classifications, so 116x6x2=1,392. The evaluation post-strata are aggregations of the 1,392 post-strata in order to provide sufficient sample size for the PES evaluation studies.

Attachment

THE FINAL 1990 PES EVALUATION POSTSTRATA

PES POSTSTRATUM	CENSUS DIVISION	PLACE TYPE	RACE/ORIGIN/TENURE GROUP
EVALUATION POSTSTRATUM #1			
210011	MIDDLE ATLANTIC	ALL C. CITIES IN TYPE I MSAs (EXC.NYC)	BLACK RENTER
210012	MIDDLE ATLANTIC	ALL C. CITIES IN TYPE I MSAs (EXC.NYC)	BLACK OWNER
220010	MIDDLE ATLANTIC	ALL CENTRAL CITIES IN TYPE II MSAs	BLACK
200011	MIDDLE ATLANTIC	CENTRAL CITIES IN NEW YORK CITY PMSA	BLACK RENTER
230020	MIDDLE ATLANTIC	ALL C. CITIES IN TYPE I MSAs (EXC.NYC)	NONBLACK HISPANIC
	MIDDLE ATLANTIC	ALL CENTRAL CITIES IN TYPE II MSAs	NONBLACK HISPANIC
200012	MIDDLE ATLANTIC	CENTRAL CITIES IN NEW YORK CITY PMSA	BLACK OWNER
200020	MIDDLE ATLANTIC	CENTRAL CITIES IN NEW YORK CITY PMSA	NONBLACK HISPANIC
200040	MIDDLE ATLANTIC	NEW YORK CITY	ASIAN & PACIFIC ISLANDER
EVALUATION POSTSTRATUM #2			
200031	MIDDLE ATLANTIC	CENTRAL CITIES IN NEW YORK CITY PMSA	ALL OTHER RENTER
210031	MIDDLE ATLANTIC	ALL C. CITIES IN TYPE I MSAs (EXC.NYC)	ALL OTHER RENTER
200032	MIDDLE ATLANTIC	CENTRAL CITIES IN NEW YORK CITY PMSA	ALL OTHER OWNER
130030	NEW ENGLAND	ALL MSA CENTRAL CITIES	ALL OTHER
210032	MIDDLE ATLANTIC	ALL C. CITIES IN TYPE I MSAs (EXC.NYC)	ALL OTHER OWNER
220030	MIDDLE ATLANTIC	ALL CENTRAL CITIES IN TYPE II MSAs	ALL OTHER
EVALUATION POSTSTRATUM #3			
245050	MIDDLE ATLANTIC	IN NEW YORK CITY PMSA, NOT IN A C. CITY	BLACK, HISPANIC
	MIDDLE ATLANTIC	IN TYPE I MSAs (EXC.NYC), NOT IN A C.C.	BLACK, HISPANIC
137050	NEW ENGLAND	ALL MSA CENTRAL CITIES	BLACK, HISPANIC
	NEW ENGLAND	MSA AREAS NOT IN CENTRAL CITIES	BLACK, HISPANIC
299950	MIDDLE ATLANTIC	IN TYPE II MSAs, NOT IN A CENTRAL CITY	BLACK, HISPANIC
	MIDDLE ATLANTIC	OTHER NON-MSA AREAS	BLACK, HISPANIC
	NEW ENGLAND	NON-MSA INCORPORATED PLACES 10,000+ POP.	BLACK, HISPANIC
	NEW ENGLAND	OTHER NON-MSA AREAS	BLACK, HISPANIC
	MIDDLE ATLANTIC	NON-MSA INCORPORATED PLACES 10,000+ POP.	BLACK, HISPANIC
360010	SOUTH ATLANTIC	IN TYPE II MSAs, NOT IN A CENTRAL CITY	BLACK
350010	SOUTH ATLANTIC	IN TYPE I MSAs, NOT IN A CENTRAL CITY	BLACK
380010	SOUTH ATLANTIC	NON-MSA INCORPORATED PLACES 10,000+ POP.	BLACK
390010	SOUTH ATLANTIC	OTHER NON-MSA AREAS	BLACK
378920	SOUTH ATLANTIC	IN TYPE I MSAs, NOT IN A CENTRAL CITY	NONBLACK HISPANIC
	SOUTH ATLANTIC	NON-MSA INCORPORATED PLACES 10,000+ POP.	NONBLACK HISPANIC
	SOUTH ATLANTIC	OTHER NON-MSA AREAS	NONBLACK HISPANIC
	SOUTH ATLANTIC	IN TYPE II MSAs, NOT IN A CENTRAL CITY	NONBLACK HISPANIC

478950	EAST SOUTH CENTRAL	NON-MSA INCORPORATED PLACES 10,000+ POP.	BLACK, HISPANIC
	EAST SOUTH CENTRAL	MSA AREAS NOT IN A CENTRAL CITY	BLACK, HISPANIC
	EAST SOUTH CENTRAL	OTHER NON-MSA AREAS	BLACK, HISPANIC
578910	WEST SOUTH CENTRAL	OTHER NON-MSA AREAS	BLACK
	WEST SOUTH CENTRAL	NON-MSA INCORPORATED PLACES 10,000+ POP.	BLACK
	WEST SOUTH CENTRAL	MSA AREAS NOT IN A CENTRAL CITY	BLACK
578920	WEST SOUTH CENTRAL	OTHER NON-MSA AREAS	NONBLACK HISPANIC
	WEST SOUTH CENTRAL	MSA AREAS NOT IN A CENTRAL CITY	NONBLACK HISPANIC
	WEST SOUTH CENTRAL	NON-MSA INCORPORATED PLACES 10,000+ POP.	NONBLACK HISPANIC
627050	EAST NORTH CENTRAL	ALL CENTRAL CITIES IN TYPE II MSAs	BLACK, HISPANIC
	EAST NORTH CENTRAL	IN TYPE I MSAs, NOT IN A CENTRAL CITY	BLACK, HISPANIC
	EAST NORTH CENTRAL	IN TYPE II MSAs, NOT IN A CENTRAL CITY	BLACK, HISPANIC
725050	WEST NORTH CENTRAL	ALL CENTRAL CITIES IN TYPE II MSAs	BLACK, HISPANIC
	WEST NORTH CENTRAL	IN TYPE I MSAs, NOT IN A CENTRAL CITY	BLACK, HISPANIC
799950	WEST NORTH CENTRAL	IN TYPE II MSAs, NOT IN A CENTRAL CITY	BLACK, HISPANIC
	WEST NORTH CENTRAL	NON-MSA INCORPORATED PLACES 10,000+ POP.	BLACK, HISPANIC
	WEST NORTH CENTRAL	OTHER NON-MSA AREAS	BLACK, HISPANIC
	EAST NORTH CENTRAL	NON-MSA INCORPORATED PLACES 10,000+ POP.	BLACK, HISPANIC
	EAST NORTH CENTRAL	OTHER NON-MSA AREAS	BLACK, HISPANIC
970010	PACIFIC	IN TYPE I MSAs, NOT IN A CENTRAL CITY	BLACK
	PACIFIC	IN TYPE II MSAs, NOT IN A CENTRAL CITY	BLACK
970020	PACIFIC	IN TYPE I MSAs, NOT IN A CENTRAL CITY	NONBLACK HISPANIC
	PACIFIC	IN TYPE II MSAs, NOT IN A CENTRAL CITY	NONBLACK HISPANIC
899950	PACIFIC	NON-MSA INC. PL.,10,000+ POP(CDPs IN HI)	BLACK, HISPANIC
	PACIFIC	OTHER NON-MSA AREAS	BLACK, HISPANIC
	MOUNTAIN	MSA AREAS NOT IN CENTRAL CITIES	BLACK, HISPANIC
	MOUNTAIN	NON-MSA INCORPORATED PLACES 10,000+ POP.	BLACK, HISPANIC
	MOUNTAIN	OTHER NON-MSA AREAS	BLACK, HISPANIC

EVALUATION POSTSTRATUM #4

260030	MIDDLE ATLANTIC	IN TYPE II MSAs, NOT IN A CENTRAL CITY	ALL OTHER
170030	NEW ENGLAND	MSA AREAS NOT IN CENTRAL CITIES	ALL OTHER
240030	MIDDLE ATLANTIC	IN NEW YORK CITY PMSA, NOT IN A C. CITY	ALL OTHER
250030	MIDDLE ATLANTIC	IN TYPE I MSAs (EXC.NYC), NOT IN A C.C.	ALL OTHER
280030	MIDDLE ATLANTIC	NON-MSA INCORPORATED PLACES 10,000+ POP.	ALL OTHER
180030	NEW ENGLAND	NON-MSA INCORPORATED PLACES 10,000+ POP.	ALL OTHER
290030	MIDDLE ATLANTIC	OTHER NON-MSA AREAS	ALL OTHER
190030	NEW ENGLAND	OTHER NON-MSA AREAS	ALL OTHER

EVALUATION POSTSTRATUM #5

500020	WEST SOUTH CENTRAL	C.C.s IN HOUSTON, DALLAS, FT.WORTH PMSAs	NONBLACK HISPANIC
500010	WEST SOUTH CENTRAL	C.C.s IN HOUSTON, DALLAS, FT.WORTH PMSAs	BLACK
310011	SOUTH ATLANTIC	ALL CENTRAL CITIES IN TYPE I MSAs	BLACK RENTER
330020	SOUTH ATLANTIC	ALL CENTRAL CITIES IN TYPE I MSAs	NONBLACK HISPANIC
	SOUTH ATLANTIC	ALL CENTRAL CITIES IN TYPE II MSAs	NONBLACK HISPANIC
430050	EAST SOUTH CENTRAL	ALL CENTRAL CITIES IN TYPE I MSAs	BLACK, HISPANIC
	EAST SOUTH CENTRAL	ALL CENTRAL CITIES IN TYPE II MSAs	BLACK, HISPANIC
310012	SOUTH ATLANTIC	ALL CENTRAL CITIES IN TYPE I MSAs	BLACK OWNER
320010	SOUTH ATLANTIC	ALL CENTRAL CITIES IN TYPE II MSAs	BLACK
530020	WEST SOUTH CENTRAL	ALL CENTRAL CITIES IN OTHER TYPE I MSAs	NONBLACK HISPANIC
	WEST SOUTH CENTRAL	ALL CENTRAL CITIES IN TYPE II MSAs	NONBLACK HISPANIC
530010	WEST SOUTH CENTRAL	ALL CENTRAL CITIES IN OTHER TYPE I MSAs	BLACK
	WEST SOUTH CENTRAL	ALL CENTRAL CITIES IN TYPE II MSAs	BLACK

EVALUATION POSTSTRATUM #6

410031	EAST SOUTH CENTRAL	ALL CENTRAL CITIES IN TYPE I MSAs	ALL OTHER RENTER
410032	EAST SOUTH CENTRAL	ALL CENTRAL CITIES IN TYPE I MSAs	ALL OTHER OWNER
500031	WEST SOUTH CENTRAL	C.C.s IN HOUSTON, DALLAS, FT.WORTH PMSAs	ALL OTHER RENTER
310031	SOUTH ATLANTIC	ALL CENTRAL CITIES IN TYPE I MSAs	ALL OTHER RENTER
500032	WEST SOUTH CENTRAL	C.C.s IN HOUSTON, DALLAS, FT.WORTH PMSAs	ALL OTHER OWNER
310032	SOUTH ATLANTIC	ALL CENTRAL CITIES IN TYPE I MSAs	ALL OTHER OWNER
510031	WEST SOUTH CENTRAL	ALL C. CITIES IN OTHER TYPE I MSAs	ALL OTHER RENTER
320030	SOUTH ATLANTIC	ALL CENTRAL CITIES IN TYPE II MSAs	ALL OTHER
520030	WEST SOUTH CENTRAL	ALL CENTRAL CITIES IN TYPE II MSAs	ALL OTHER
510032	WEST SOUTH CENTRAL	ALL C. CITIES IN OTHER TYPE I MSAs	ALL OTHER OWNER
420030	EAST SOUTH CENTRAL	ALL CENTRAL CITIES IN TYPE II MSAs	ALL OTHER

EVALUATION POSTSTRATUM #7

360030	SOUTH ATLANTIC	IN TYPE II MSAs, NOT IN A CENTRAL CITY	ALL OTHER
570030	WEST SOUTH CENTRAL	MSA AREAS NOT IN A CENTRAL CITY	ALL OTHER
350030	SOUTH ATLANTIC	IN TYPE I MSAs, NOT IN A CENTRAL CITY	ALL OTHER
470030	EAST SOUTH CENTRAL	MSA AREAS NOT IN A CENTRAL CITY	ALL OTHER
380030	SOUTH ATLANTIC	NON-MSA INCORPORATED PLACES 10,000+ POP.	ALL OTHER
580030	WEST SOUTH CENTRAL	NON-MSA INCORPORATED PLACES 10,000+ POP.	ALL OTHER
480030	EAST SOUTH CENTRAL	NON-MSA INCORPORATED PLACES 10,000+ POP.	ALL OTHER
590030	WEST SOUTH CENTRAL	OTHER NON-MSA AREAS	ALL OTHER
390030	SOUTH ATLANTIC	OTHER NON-MSA AREAS	ALL OTHER
490030	EAST SOUTH CENTRAL	OTHER NON-MSA AREAS	ALL OTHER

EVALUATION POSTSTRATUM #8

601020	EAST NORTH CENTRAL	C. CITIES IN CHICAGO, DETROIT PMSAs	NONBLACK HISPANIC
	EAST NORTH CENTRAL	ALL C. CITIES IN OTHER TYPE I MSAs	NONBLACK HISPANIC
600011	EAST NORTH CENTRAL	C. CITIES IN CHICAGO, DETROIT PMSAs	BLACK RENTER
600012	EAST NORTH CENTRAL	C. CITIES IN CHICAGO, DETROIT PMSAs	BLACK OWNER
710050	WEST NORTH CENTRAL	ALL CENTRAL CITIES IN TYPE I MSAs	BLACK, HISPANIC
610011	EAST NORTH CENTRAL	ALL C. CITIES IN OTHER TYPE I MSAs	BLACK RENTER
610012	EAST NORTH CENTRAL	ALL C. CITIES IN OTHER TYPE I MSAs	BLACK OWNER

EVALUATION POSTSTRATUM #9

600031	EAST NORTH CENTRAL	C. CITIES IN CHICAGO, DETROIT PMSAs	ALL OTHER RENTER
600032	EAST NORTH CENTRAL	C. CITIES IN CHICAGO, DETROIT PMSAs	ALL OTHER OWNER
710031	WEST NORTH CENTRAL	ALL CENTRAL CITIES IN TYPE I MSAs	ALL OTHER RENTER
610031	EAST NORTH CENTRAL	ALL C. CITIES IN OTHER TYPE I MSAs	ALL OTHER RENTER
710032	WEST NORTH CENTRAL	ALL CENTRAL CITIES IN TYPE I MSAs	ALL OTHER OWNER
620030	EAST NORTH CENTRAL	ALL CENTRAL CITIES IN TYPE II MSAs	ALL OTHER
610032	EAST NORTH CENTRAL	ALL C. CITIES IN OTHER TYPE I MSAs	ALL OTHER OWNER
720030	WEST NORTH CENTRAL	ALL CENTRAL CITIES IN TYPE II MSAs	ALL OTHER

EVALUATION POSTSTRATUM #10

650030	EAST NORTH CENTRAL	IN TYPE I MSAs, NOT IN A CENTRAL CITY	ALL OTHER
660030	EAST NORTH CENTRAL	IN TYPE II MSAs, NOT IN A CENTRAL CITY	ALL OTHER
750030	WEST NORTH CENTRAL	IN TYPE I MSAs, NOT IN A CENTRAL CITY	ALL OTHER
760030	WEST NORTH CENTRAL	IN TYPE II MSAs, NOT IN A CENTRAL CITY	ALL OTHER
680030	EAST NORTH CENTRAL	NON-MSA INCORPORATED PLACES 10,000+ POP.	ALL OTHER
780030	WEST NORTH CENTRAL	NON-MSA INCORPORATED PLACES 10,000+ POP.	ALL OTHER
690030	EAST NORTH CENTRAL	OTHER NON-MSA AREAS	ALL OTHER
790030	WEST NORTH CENTRAL	OTHER NON-MSA AREAS	ALL OTHER

EVALUATION POSTSTRATUM #11

903011	PACIFIC	C. CITIES IN LOS ANGELES-LONG BEACH PMSA	BLACK RENTER
	PACIFIC	ALL CENTRAL CITIES IN OTHER TYPE I MSAs	BLACK RENTER
	PACIFIC	ALL CENTRAL CITIES IN TYPE II MSAs	BLACK RENTER
903021	PACIFIC	C. CITIES IN LOS ANGELES-LONG BEACH PMSA	NONBLACK HISPANIC, RENTER
	PACIFIC	ALL CENTRAL CITIES IN OTHER TYPE I MSAs	NONBLACK HISPANIC, RENTER
	PACIFIC	ALL CENTRAL CITIES IN TYPE II MSAs	NONBLACK HISPANIC, RENTER
830050	MOUNTAIN	ALL CENTRAL CITIES IN TYPE I MSAs	BLACK, HISPANIC
	MOUNTAIN	ALL CENTRAL CITIES IN TYPE II MSAs	BLACK, HISPANIC
903012	PACIFIC	C. CITIES IN LOS ANGELES-LONG BEACH PMSA	BLACK OWNER
	PACIFIC	ALL CENTRAL CITIES IN OTHER TYPE I MSAs	BLACK OWNER
	PACIFIC	ALL CENTRAL CITIES IN TYPE II MSAs	BLACK OWNER
903022	PACIFIC	C. CITIES IN LOS ANGELES-LONG BEACH PMSA	NONBLACK HISPANIC, OWNER
	PACIFIC	ALL CENTRAL CITIES IN OTHER TYPE I MSAs	NONBLACK HISPANIC, OWNER
	PACIFIC	ALL CENTRAL CITIES IN TYPE II MSAs	NONBLACK HISPANIC, OWNER
903041	PACIFIC	C. CITIES OF ALL MSAS	ASIAN & PACIFIC ISLANDER, RENTER
903042	PACIFIC	C. CITIES OF ALL MSAS	ASIAN & PACIFIC ISLANDER, OWNER

EVALUATION POSTSTRATUM #12

910031	PACIFIC	ALL CENTRAL CITIES IN OTHER TYPE I MSAs	ALL OTHER RENTER
900031	PACIFIC	C. CITIES IN LOS ANGELES-LONG BEACH PMSA	ALL OTHER RENTER
810031	MOUNTAIN	ALL CENTRAL CITIES IN TYPE I MSAs	ALL OTHER RENTER
910032	PACIFIC	ALL CENTRAL CITIES IN OTHER TYPE I MSAs	ALL OTHER OWNER
900032	PACIFIC	C. CITIES IN LOS ANGELES-LONG BEACH PMSA	ALL OTHER OWNER
810032	MOUNTAIN	ALL CENTRAL CITIES IN TYPE I MSAs	ALL OTHER OWNER
920030	PACIFIC	ALL CENTRAL CITIES IN TYPE II MSAs	ALL OTHER
820030	MOUNTAIN	ALL CENTRAL CITIES IN TYPE II MSAs	ALL OTHER

EVALUATION POSTSTRATUM #13

960030	PACIFIC	IN TYPE II MSAs, NOT IN A CENTRAL CITY	ALL OTHER
870030	MOUNTAIN	MSA AREAS NOT IN A CENTRAL CITY	ALL OTHER
950030	PACIFIC	IN TYPE I MSAs, NOT IN A CENTRAL CITY	ALL OTHER
880030	MOUNTAIN	NON-MSA INCORPORATED PLACES 10,000+ POP.	ALL OTHER
980030	PACIFIC	NON-MSA INC. PL.,10,000+ POP(CDPs IN HI)	ALL OTHER
890030	MOUNTAIN	OTHER NON-MSA AREAS	ALL OTHER
990030	PACIFIC	OTHER NON-MSA AREAS	ALL OTHER
970040	PACIFIC	MSA AREAS NOT IN A CENTRAL CITY	ASIAN & PACIFIC ISLANDER
989040	PACIFIC	NON-MSA AREAS	ASIAN & PACIFIC ISLANDER
090060	ALL	AM.IND. IN AM. IND. SAMPLING STRATA	ALL

APPENDIX 6: TECHNICAL DISCUSSION OF STATISTICAL ISSUES RELEVANT TO ADJUSTMENT

This appendix is a technical supplement to the general findings of the Undercount Steering Committee. The report by the full committee captures the general sense of the committee's deliberation in a manner that should be accessible to a fairly wide audience. The following commentary is not constrained by a similar attempt at accessibility. It describes several of the most important theoretical considerations underlying the technical recommendation and the limits of knowledge at this date.

In doing so, the appendix serves two purposes. First, it describes technical issues that have occupied the attention of committee members with the most extensive background in mathematical statistics. Second, it states technical opinions, judgments and levels of uncertainty about these issues. It attempts to distinguish instances where the technical evidence for specific views is relatively strong and those where it is weak. The discussion will be deliberately unbalanced: it focuses primarily on issues that give the greatest technical concern about the general findings of the committee.

For the nontechnical reader, it is important to place these concerns within an appropriate context. There are things that are right about the production adjustments, particularly for high levels of aggregation. The concerns presented in this appendix do not in themselves nullify concurrence with the recommendation of the majority in the assessment that the adjustment would make an overall improvement to the census counts for States and large areas. At the same time, members' judgment depends to some degree on the face-validity of the adjustments. The preference of the statistical experts on the USC would have been for an assessment where face-validity complemented a complete and extensive technical evaluation.

The issues that are of greatest statistical concern with respect to the adjustment:

1. The underestimation of uncertainty for the smoothing model.

2. The likely case that the loss function analysis presented to the committee on or before June 14 contained choices that indirectly favored adjustment.

3. The incompleteness of the loss function analysis to examine alternative assumptions, including investigating the effects of concerns expressed by some committee members.

4. The limited time to examine a set of secondary but important issues, e.g.,

    a. some concerns for nonsampling errors where there has been insufficient time to investigate, specifically:

        i. Reexamination of the evaluation studies/total error model along dimensions that are important to adjustment, e.g., owner/renter, allocations.

b. Evidence on the effect of adjustment on other uses of the census:

i. The count and geographic distribution of the population 65+.

ii. Likelihood that within-state distributions may be harmed by adjustment in a minority of States, primarily or exclusively those with low minority populations.

1. Understatement of uncertainty by the smoothing model. In the notation of a March 7, 1989 letter from Wayne A. Fuller to Cary Isaki, the basis for the smoothing is the model:

$$Y = X\beta + w + e, \tag{1}$$

where $y = X\beta + w$ is the true value of the adjustment factors to be predicted, and it is assumed that

$$(w', e')' \sim N(0, \text{block diag}(I\sigma^2, V)),$$

where V, representing the variance-covariance matrix of the sampling error terms of the sampling errors e, is assumed known, and σ^2, representing the variance of the individual components of the random effects, w, is to be estimated. Fuller writes the predictor as

$$\tilde{y} = Y - V\hat{\Sigma}_{zz}^{-1}(Y - X\hat{\beta}), \tag{2}$$

where $Z = w + e$, and

$$\hat{\Sigma}_{zz} = I\hat{\sigma}^2 + V.$$

Fuller expresses the error of prediction as

$$\begin{aligned} \tilde{y} - y &= e - V\hat{\Sigma}_{zz}^{-1}(w + e) \\ &\quad + V\hat{\Sigma}_{zz}^{-1}X(\hat{\beta} - \beta) \\ &\quad - V[\hat{\Sigma}_{zz}^{-1} - \Sigma_{zz}^{-1}](Y - X\beta) \\ &\quad + O_p(n^{-1}). \end{aligned} \tag{3}$$

Equation (3) groups the right-hand side into three terms, each on one of the first three lines, and a lower-order residual term on the fourth line that is dropped from the subsequent approximations. Fuller expresses the estimated variance of the term on the first line as

$$V - V\hat{\Sigma}_{zz}^{-1}V. \tag{4}$$

His expression for the variance of the term on the second line of (3) is

$$V\Sigma_{zz}^{-1}X[\hat{V}(\hat{\beta})]X^T\hat{\Sigma}_{zz}^{-1}V. \tag{5}$$

which depends on the estimated variance of the regression coefficients, $\hat{V}(\hat{\beta})$. To apply (5), the production estimates have employed

$$\hat{V}(\hat{\beta}) = (X^T\hat{\Sigma}_{zz}^{-1}X)^{-1}. \tag{6}$$

The approximate variance of the third line of (3) is

$$E\{V\Sigma_{zz}^{-1}(\hat{\Sigma}_{zz} - \Sigma_{zz})\Sigma_{zz}^{-1}(Y - X\beta)(Y - X\beta)^T\Sigma_{zz}^{-1}(\hat{\Sigma}_{zz} - \Sigma_{zz})\Sigma_{zz}^{-1}V\}, \tag{7}$$

which Fuller estimates by

$$V\hat{\Sigma}_{zz}^{-1}[\hat{\Sigma}_{zz}^{-1}(Y - X\hat{\beta})(Y - X\hat{\beta})^T\Sigma_{zz}^{-1}]\Sigma_{zz}^{-1}V(\hat{V}\{\hat{\sigma}^2\}). \tag{8}$$

where $\hat{V}\{\hat{\sigma}^2\}$ is the estimated variance of $\hat{\sigma}^2$. In the production smoothing, $\hat{V}\{\hat{\sigma}^2\}$ was estimated indirectly by examining variation in the log-likelihood function as a function of σ^2.

Although not obvious from a cursory inspection, covariances among the three major terms of (3) vanish under the assumptions of the model.

The variance estimator was subsequently revised to take into account the application of ratio estimation at the regional level. Although the revision added terms to the expansion, fundamental issues in the appropriateness of the variance estimation are more easily discussed without considering this complication.

The three terms, (4), (5), and (8), in the variance estimator correspond to distinct components of the uncertainty. The most important term in the production estimates is (5), which represents the uncertainty due to estimation of the regression coefficients from the data. In the instance that $\sigma^2=0$, only (5) contributes to the uncertainty, and it simplifies to the classical expression for the variance of the predicted values in the usual regression problem with general, although nonsingular, covariances. For $\sigma^2>0$, (4) captures two components of the variance: 1) the variance from the composite estimate (2) that combines both the regression prediction and Y, so that the effect of the sampling variance of Y must be included; and 2) the uncertainty associated with the random effects w. Finally, (8) attempts to reflect the additional variance from estimating σ^2 from the data instead of knowing its true value.

It is likely that the estimation of uncertainty for the smoothed estimates, although formally correct to the order of approximation under these assumptions, is a substantial understatement of the true uncertainty. More specifically, there are reasons to be particularly concerned with the validity of inferences for higher-level aggregates, such as the poststratum groups (the 116 groups formed by collapsing the poststrata across the 12 age/sex categories), rather than for the basic units of analysis for the smoothing model, the 1392 poststrata. The concern stems from several sources:

1a. No explicit allowance for the effect of selection of variables on the overall variance. Through (6), the production variance estimator assumes that the regression variables were all individually preselected, whereas in fact some were preselected while many entered the model through the prespecified variable-selection procedure. Since the variable selection procedure identified the variables that best described the observed sample estimates, the predicted values from the regression more closely resembled the sample estimates than would have a linear regression with an entirely preselected set with the same number of predictors. Consequently, the predicted values from variable-selection regression have a higher variance than the predicted values from a regression with prespecified variables, but no explicit calculation of this increase in uncertainty has been completed, primarily because no analytic expression is available.

1b. The term (8) arises from estimating the contribution to uncertainty from the estimation of σ^2 in the third line of (3), but assumes that V is fixed. The same expansion would imply additional uncertainty due to the estimation of V. Since the production model employed raw correlations between adjustment factors in estimating V, there is potentially significant contributions to the overall variance from this component. An alternative smoothing, based on a V with smoothed correlations, produced substantially different results in the two regions in which this was attempted. At this date, there has been no direct evaluation of the effect of esimating V on the uncertainty of the production estimates.

1c. The smoothing model assumes that the distribution of errors about the regression fit, w, has a covariance matrix of the form $\sigma^2 I$. There has not been sufficient time to evaluate the sensitivity of the model to alternative reasonable assumptions, which would in general allow for positive covariances of the elements of w in the model among poststrata in the same poststratum groups. For example, the deviation from the model for males ages 0-9 in a given poststratum group are likely to be correlated highly with the deviation for females 0-9 and to a lesser but still significant degree with coverage for their parents, females 20-29, 30-44, males 20-29, and 30-44. Indeed, a more reasonable assumption would be that errors are correlated among most age groups within each poststratum group. (This is the situation for the sampling variances:

a substantial proportion of the sampling variance for a poststratum group comes from correlations among the poststrata comprising the poststratum group.) In turn, these correlations would produce a higher level of uncertainty for poststratum groups and generally other higher-level aggregates, such as States.

1d. There is a potential for inefficiency in the linear regression due to a problem frequently termed the "ecological fallacy." The linear combination of predictors selected to give the best fit at the poststratum level, with the age/sex detail, may give an inefficient or biased estimates of the optimal linear fit for poststratum groups. This issue has not been directly evaluated, but it may explain some differences between the production model and the 116-point alternative.

The production model has proven too complex to evaluate directly within the available time. In hindsight, a simpler model would have facilitated the task of model evaluation. Proposed techniques for evaluating the production model directly, particularly with respect to issue 1a, involve replication of the production model multiple times, which would have required greater staff and computer resources than available. This absence of a complete direct evaluation increased the importance of the examination of alternative models, since the alternative models serve in part as indirect evaluations of these issues. For example, issue 1b was partially addressed through a procedure to smooth the correlation matrices before input to the production smoothing procedure; an outcome showing only small changes in the estimated factors would have provided evidence minimizing the importance of issue 1b. In fact, the appearance of differences on the order of one standard error from the production model values points to significant components of variance currently underestimated, but leaves the issue without a satisfactory quantification.

Fitting the 116-point model represented in part an attempt to assess 1a, 1c, and 1d simultaneously, although in each case indirectly. The structure of the 116-point model (actually, separate models for 49 minority points and 66 non-minority points, without reevaluating the model for American Indians) included a more limited set of variables selected through a step-wise procedure at a national level, resulting in far fewer predictors. Although the 116-point alternatives were not entirely free from selection effects, the comparison to the production model affords some notion of the effect of model selection, thus addressing 1a. By expressing the empirical Bayes model at the poststratum group level, the concerns expressed by 1c and 1d were effectively avoided by the alternative. Again, modest differences would have been reassuring, but the relatively large differences, although leaving any of the smoothing models closer to the PES than the adjusted census, leave the validity of inferences about the reliability of the smoothed estimates in some question.

Summary reaction and judgment: It is likely that 1a, 1b, and 1c add considerably to the total uncertainty. It is possible that the contribution from 1d is nontrivial as well, although none of the evidence examined so far isolates the effect of 1d as separately important. As a matter of judgment, the total understatement of variance of the estimates from the smoothing model may be in the range of a factor of 1.7 to 3.0 in terms of variance, or 1.3 to 1.7 on the standard error. In other words, valid confidence intervals for smoothed estimates by state may be about 1.3 to 1.7 times larger than we have stated them.

The standard errors for the unsmoothed State estimates serve as an upper bounds of sorts, for it is hard to believe that the smoothing did not at least produce some gains relative to the unsmoothed estimates.

2. The likely case that the loss function analysis presented to the committee up to June 14 contained choices that inadvertently favored adjustment. When meeting as a committee as a whole, we did discuss in a general way the question of whether the framing of the loss function analysis indirectly favored the PES over the census by design, in part through the manner in which the target population was constructed. This concern did have some basis in fact, and work proceeded to examine a correction to the problem. In short, a more appropriate comparison for a squared difference measure involves reducing the estimated risk of using the census by the PES sampling variance in estimating the target. An attempt is underway to employ the bootstrap to estimate corrections for other loss functions.

The necessity to correct the census risk for uncertainty in estimating the target in the loss function analysis now interacts with appropriate estimation of the uncertainty for the smoothing. The corrections to the loss function analysis implemented so far depend on the appropriateness of the variance-covariance matrix estimated for the smoothed factors, obtained as the sum of (4), (5) and (8), with further modifications to account for the effect of ratio estimation. If, in fact, this covariance matrix is underestimated, then there is a corresponding implication of additional uncertainty and lack of balance in the loss function analysis.

Reaction and Comment: The USC expressed a concern earlier that aspects of the loss function analysis may have favored the PES, and the committee attempted to correct judgments both for this systematic bias and for omissions from the analysis of some components of error. This was not an easy task and led to uncertainty about the outcome.

3. The incompleteness of the loss function analysis to examine alternative assumptions, including investigating the effects of concerns expressed by other committee members. At this date, the loss function analysis is set up under one set of reasonable assumptions, but there has not been time to examine sensitivity to those assumptions. Ideally, and with more time, it would be possible to enable the loss function analysis to reflect concerns expressed by a minority of the committee over components of error judged unimportant or set aside by the majority of the committee.

For example, the loss function analysis could be expanded to make allowances for small departures from the synthetic assumption that were within a confidence interval or range suggested by the data, in place of completely omitting any contribution. Analyses from P12, including some that were not completed, could have been used as a basis for alternative models for the effect of departures from the synthetic assumption.

We would also then be able to quantify how large a magnitude of error, such as unmeasured biases in the evaluation studies, would be required before the conclusions of the loss function analysis would be substantially altered.

4. The limited time to examine a set of secondary but important issues. The outline lists two examples of issues that have arisen that the committee would have preferred more time to investigate. One is that the predefined evaluation poststrata were sensible yet may not represent the most effective way to evaluate the PES. In particular, the loss function analysis is likely to more heavily penalize the PES for errors whose effects are positively correlated with the estimated undercount level in the PES compared to those errors that are more random; this would suggest, for example, the importance of reexamining the evaluation data along an owner/renter dimension, since this dimension appears strongly correlated with the estimated undercounts. The counterintuitive coefficient for census allocations obtained for the 66-point nonminority regression begs an explanation, but we have not had time to investigate this.

The outline also lists a level of secondary concerns where we would like to see more effort and consider the direct evidence close to completely lacking. For one, the large number of estimated overcounts for poststrata of those age 65+ raises a concern about whether the adjustment is moving the census figures closer towards the truth for this group. This appears to be an open question: We think the evidence has yet to be examined closely enough to provide an unambiguous answer. Similarly, an appropriate loss function analysis may show that, because of variability in estimating the adjustment factors, the expected risk for the adjusted count for uses within-State is higher than the unadjusted census for several small states with very low concentrations of minority populations. An adjustment could proceed without such information, but an expeditious and complete investigation of these issues seems desirable.

Addendum to the Report of the Undercount Steering Committee

The purpose of this addendum is to present some additional information related to the Loss Function analysis that has been developed since June 21, 1991.

The addendum has four parts:

1. Results for a revised loss function,

2. Results for a simulation of a different variance estimator for the smoothed factors,

3. Results of a model for expected number of states with losses from adjustment, in the case that the census counts and adjusted counts are equally accurate, and

4. Summary comments on how USC places the new information into the context of all the other evaluation information.

Because this document is physically separate from its full report, the USC believes that the ties to the past report must be explicit.

1. Revised Loss Function

As background, consider the squared error loss function analysis, which is based on comparing L_{ci} and L_{ai} where

$$L_{ci} = (P_{ci} - P_{ti})^2$$

$$L_{ai} = \sum_{q=1}^{1,000} \frac{(P_{aiq} - P_{ti})^2}{1,000}$$

and where

X_{ci} = the unadjusted census count for the i^{th} state (or other area as appropriate).

P_{ci} = $X_{ci}/X_{c.}$, and a "." in a subscript denotes summation over that subscript.

X_{ai}, X_{ti}, P_{aiq}, and P_{ti} are defined analogously for the adjusted census and target populations, respectively. The subscript q denotes the q^{th} draw from the adjusted census population distribution.

U_i = $P_{ci} - E(P_{ti})$ The bias in the census assuming that P_{ti} is an unbiased estimator of "truth"

B_i = $P_{ai} - E(P_{ti})$ The bias in the PES also assuming that P_{ti} is unbiased.

There are several models that involve assumptions about the relationship between P_{ti} and π_i, where π_i is the true proportion of the population in the i^{th} state (or other area as appropriate). In essence, the original model specified that the P_{ti} were fixed quantities, representing the best estimate of the π_i, and that variation in P_{ci} over realizations of the census was not relevant to the loss analysis. Under that assumption, the L_{ci} was subject to no variance; it too was a fixed quantity. This model tends to favor the adjusted census since the P_{ti} are subject to variability that inflate the estimated losses for the census.

Another model acknowledges the fact that the P_{ti} are sample-based estimates of π_i, but still maintains that variation in P_{ci} over realizations of the census can be ignored. Under this model the squared error loss function analysis is a comparison of the squared bias of the census and mean square error of the PES. Given the sample-based nature of P_{ti}, we want $E(L_{ci})$ to be equal to the squared bias of the census and $E(L_{ai})$ to be equal to the sum of the sampling variance and the bias in the PES.

Consider the expected values of L_{ci} and L_{ai}

$$E(L_{ci}) = E(P_{ci} - P_{ti})^2 = VAR(P_{ti}) + (P_{ci} - E(P_{ti}))^2$$

$$= VAR(P_{ai}) + VAR(\hat{B}_i) + U_i^2$$

$$E(L_{ai}) = E\left(\sum_{q=1}^{1,000} \frac{(P_{aiq} - P_{ti})^2}{1,000}\right)$$

$$= VAR(P_{ai}) + VAR(\hat{B}_i) + B_i^2$$

(Note that we are assuming that the $Cov(P_{ai},B_i)$ is negligible in deriving the above expression.)

The value of $E(L_{ci})$ is biased upwards by $Var(P_{ai}) + Var(B_i)$. The loss associated with the census is inflated incorrectly by the variance of the PES sampling. The loss function analysis has been partially corrected for the squared error loss by using the following estimator for L_{ci}

$$L_{ci}^{*} = (P_{ci} - P_{ti})^2 - V\hat{A}R(P_{ai})$$

The expected value of L_{ci}^{*} is

$$E(L_{ci}^{*}) = U_i^2 + VAR(\hat{B}_i)$$

The additional error due to $Var(B_i)$ will not be of concern since it cancels in the difference $L_{ci} - L_{ai}$.

This work has been carried out and is documented in a separate memorandum to the Undercount Steering Committee. This new analysis shows that 21 states are made less accurate from adjustment. (Note that the number 21 itself is an estimate based on the given PES sample, and we cannot be sure that the individual States indicated as harmed by the analysis would in fact be the ones harmed on the average over repetitions of the PES sample.) However, the loss for the unadjusted census is still considerably larger than the loss for the adjusted census; that is, the total squared loss is about 72.4×10^{-7} for the census and about 7×10^{-7} for the adjusted census.

We note that other models are theoretically justifiable for the loss function; for example, acknowledging model-selection variance in P_{ai} and P_{ti}, and variation in realizations of the census, in P_{ci}. Finally, informal analysis of the adjusted census data indicates that an analogous loss function analysis carried out for population totals (instead of for proportions) would be more favorable to the adjusted census.

2. Reflecting Model Variance in P_{ai}

There are also concerns that the variance of P_{ai} may be understated. A sensitivity analysis motivated by these concerns was carried out where the variance of P_{ai} was first increased by 50 percent and then doubled. The number of States made less accurate from adjustment increased to 27 and 29 respectively for the two sensitivity treatments.

On the other hand, in one respect the correction to the loss function was probably an overcorrection, because the estimates of the variances used were based on calculations omitting a covariance term. Revised estimates of the variance, based on the loss function simulations themselves, imply a lower correction. A simplified reanalysis gives an estimate of 18 States with negative signs, instead of 21. Increasing the simulation variance by 50 percent brings the number to 22; a doubling produces 28.

3. Expected Number of States with Losses from Adjustment, under Equal Accuracy of Census Counts and Adjusted Counts

The 21 negative differences obtained from the revised loss function analysis for States, because it is so close to half of the States, creates the impression that the observed result is close to a break-even point (where adjustment does as much harm as good). Further calculations indicate that the result is not as negative towards the adjusted census as it might at first seem.

Suppose that the true population proportion for state i, π_i, is equal to the census proportion, P_{ci} modified by an error w_i. For simplicity, assume that $w_i \sim N(0, P_{ci}^2 * .01^2)$, i.e., with a standard deviation of 1 percent of the census proportion. Further, suppose that the PES production estimate is an unbiased estimate of π_i with a standard deviation of 1 percent of the census proportion. In other words, the census distribution is a close approximation to the truth but differs randomly by about 1 percent, and the PES is unbiased but has a sampling variance of the same magnitude. Thus, the sets of two proportions are equally accurate.

We have:

$$P_{ci} = P_{ti} - w_i$$

$$P_{ai} = P_{ti} + e_i$$

For simplicity, we ignore the constraints and assume that the w_i's and e_i's are independent.

The loss function analysis includes corrections for estimated bias in the PES, but suppose, again for simplicity, that the correction may be ignored. The loss function analysis compares

$$L_{ci} = (e_i + w_i)^2 - Var(e_i)$$

to

$$L_{ai} = Var(e_i).$$

Thus a negative sign for the difference occurs for any state with estimates with

$$(e_i+w_i)^2 < 2\,Var(e_i)$$

The left-hand side is the variance of $w_i + e_i$. Under assumptions of normality, the negative sign should appear in an expected 68 percent of the states, or about 34 out of 50. Intuition that the break-even point is when half of the states have negative losses and half have positive is <u>not</u> correct.

Further, when the ratio of var(w_i) to var(e_i) is 2 to 1, strongly favoring an adjustment, then a similar calculation gives that 59 percent would have negative signs, or about 29.5 States. The expected number of negative signs is about 21 at a ratio of var(w_i) to var(e_i) of 5 to 1. Under this simple model, observing 21 negative signs is consistent with a strong positive effect of adjustment on the measurement of the true population proportions.

Inspection of the State-level data indicates a high degree of clustering of the results by division. This clustering does not substantially affect validity of the proportions computed under this model, but it prevents valid inferences from the number of original conclusion, while acknowledging that the ongoing work, had it been available by the date our recommendation was due, may have caused different "weighting" of the results.