STATEMENT OF JERRY L. COFFEY, Ph.D.,
MATHEMATICAL STATISTICIAN
BEFORE THE
SUBCOMMITTEE ON THE CENSUS
COMMITTEE ON GOVERNMENT REFORM AND OVERSIGHT
U.S. HOUSE OF REPRESENTATIVES
MAY 5, 1998

Thank you, Mr. Chairman and members of the Subcommittee for the opportunity to comment on these important census issues. I would also like to thank the Subcommittee Staff for the many documents they provided on the issues being discussed today, especially the extraordinary papers from Kenneth Darga who just spoke to you. I retired from the Statistical Policy office in OMB last year. I don't miss the daily flood of tough statistical issues that came across my desk, but I did find it stimulating to get a fresh look at this set of issues so important to the plan for Census 2000.

I would like to comment on two related issues. The first arises from a remarkable report generated by senior Census Bureau staff and a panel of experts called the "Report of the Committee on Adjustment of Postcensal Estimates." This analysis was undertaken after the Secretary of Commerce had made the adjustment decision when he asked the Census Bureau to examine the suitability of the adjustment methodology known as Dual System Estimation (DSE) for making intercensal estimates.

The evaluation described in the CAPE report is one of those rare opportunities to back away from a politically charged issue and examine facts and methods as objectively as possible. I have some reservations about the ground rules of the study, but I have worked with many of the CAPE committee members and some members of the Expert Panel over the years and I know them to be individuals of extraordinary professional competence and integrity.

Conceptually, the Dual System Estimation approach looked at the records collected in the sample areas covered by the Post Enumeration Survey (PES) and logically defined four "cells." The largest cell consisted of those cases where records from the PES and the actual enumeration were able to be matched. Cells two and three consisted of records that appeared in either the PES or the actual enumeration with no apparent matching record in the other data set. The fourth cell consists of those hypothetical cases for which there are no records in either data set -- households or persons who were never found by the actual enumeration nor the PES -- I use the term hypothetical to emphasize the fact that there is no data for these cases.

For the first three cells a statistician has three things to work with -- expectations (based on prior experience and assumptions), actual data from one data set or both, and performance information. For the fourth cell, there is no data and thus no performance information, so you are left with only expectations. While expectations, data and performance information are all potential sources of error, there is a lot more you can do with data and performance information than without it.

The Committee actually did quite a lot with this information. It found some simple performance errors that exaggerated the original PES estimates of "undercount" by about 19% (.40 out of 2.08).

It then found that about 45% of what was left (.73 out of 1.58) was attributable to measurable bias. The report put it in even stronger terms --

> "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, 50% or more of the estimated undercount is bias." (CAPE report, p.15)

The first bias was removable (and was removed in the revised estimate). The Census Bureau's expert panel urged them to attempt to remove the second (larger) bias, but the Bureau determined that it could not be removed without risking even larger errors. So after all the years of planning that produced the original estimate and another year of intensive evaluation of that estimate, it turned out to be about 40% measured undercount and 60% measured bias. (An independent 1994 analysis of the PES cited in Mr. Darga's paper put these figures at 30% and 70%, respectively.) At this stage of the evaluation, the expert panel and the committee were asking the question statisticians should always ask -- "Are we measuring what we think we are measuring?" The likelihood that a minor fraction of what was measured looked like undercount produced considerable discomfort and the fact that the bias was inextricably interwoven with the apparent undercount effects made matters worse.

In theory there can be offsetting unmeasurable bias, e.g., the "correlation bias." This form of bias involves assumptions that are not strictly satisfied and the size of that fourth cell where there is no data or performance information to analyze because households or individuals were missed by both systems. But if you think about this situation you can begin to see a "Catch 22."

For Dual System Estimation to work, the unobserved fourth cell must be small. If both the actual enumeration and the later sample miss a substantial proportion of the uncounted population, then the DSE process unravels, since the attributes of the measured portion can no longer be attributed to the whole uncounted group without substantial risk of (additional) bias.

On the other hand, if the fourth cell is small, then the offsetting bias is small and one is left with a measured undercount about half the size implied by Demographic Analysis. Demographic Analysis is not perfect (in fact Attachment 8 of the Cape report notes some of the sources of error), but few would be comfortable with the idea that it missed the undercount estimate by a factor of two or more.

For a statistician there is a third problem. If the DSE methodology choked on its own error trying to measure a 2% phenomenon, that amount of error will completely overwhelm a phenomenon as small as ½ or 1%.

In the report, the Census Bureau assumes a moderately small correlation bias which did not fully offset the measured bias and thus was equivalent to a measured undercount of about 1.2% (net of bias). The remainder of the analysis was an intense and conscientious search for an internally consistent interpretation of the assumptions and the facts. Each different chain of reasoning was pursued until all of them ultimately foundered on results that were inconclusive or even impossible (e.g., negative values in the fourth cell).

The report summarizes a comprehensive and thoroughly professional review of the 1990 experience with DSE, but it doesn't end there. In the last few sections it charts the trail of the findings from the Expert Panel to the Census Bureau Committee and finally to the Director and the executive staff. Along this route one other interesting thing happened.

> "Prior to the main part of the [CAPE decision discussion] 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." (p.31, CAPE report)

After all of the research and analysis that had been done, this simple "split-the-difference" idea didn't sit well with some members of the Committee. On the other hand --

> "Analysis done by [the] 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." (p.32, CAPE report)

In hindsight, this result is not that surprising. There is a very powerful theory, called Bayes theory, that sometimes produces simple estimators like this one and I suspect the proponent of the composite approach may have been a closet Bayesian. In this case the composite estimate is logically equivalent to cutting the adjustment increment due to Dual System Estimation in half, and since about half that increment is bias, the result is not surprising.

The bottom line is -- this was an extraordinary effort in a politically neutral environment, "based on a massive amount of data" and a "large volume of additional research." But in the end it left profound doubts about the DSE methodology. As the report put it (somewhat delicately) on page 31 --

> "Below the State level the Committee could not make a recommendation about improvement from adjustment and supported the census counts." (p.31, CAPE report)

The last page of the report voiced specific "concerns about the future raised during final discussions." The very first point urged the Census Bureau to "examine alternatives to the Dual System Estimation process..." since "Some of the problems of this approach may continue despite best efforts, meaning that a full adjustment based on such a system might never be possible." (p.34, CAPE report)

There was a prescient comment from a member of the Expert Panel concerning the year of work that occurred after the official undercount estimate had been corrected downward by 20% --

> "He cautioned that if new research between July 1991 and [July 1992] uncovered new findings, he would not be surprised to see additional research after July 1992 turn up new results and new estimates of undercount." (Attachment 8 to the CAPE report)

Now it is eight years after the 1990 Census and researchers are still finding significant new problems. By mid-1992, about half of the DSE estimate was attributable to measured bias. The later research

cited in Mr. Darga's paper raised the figure to 70% of the result attributable to measurement error. I understand that there is a later estimate that puts the DSE estimate of undercount at 20% undercount and 80% error, but I have not seen that paper. Anyone who is concerned about the accuracy of the Census must read and understand the papers prepared by Kenneth Darga.

These problems are not going to be solved by increasing sample size. Bias is not so well-behaved as sampling error -- it is notorious for persisting undiminished even when sampling error is reduced. This brings me to my second concern about the proposed plan for the 2000 Census -- the decision not to pursue the last 10% of the countable population during the actual enumeration.

In my 32 years as a government statistician, I have never found anyone willing to argue that truncating follow-up will improve the quality of the data. The excuses have always been about saving money or saving time -- and this case is no different. After a very brief period allowed for response to the initial mail-out, the plan proposes to randomly eliminate about 10% of the population from the intensive follow-up operations that have been used in prior censuses. In essence this decision expands the uncounted portion of the population by a factor of five or more.

This is not to say that some things won't look better. The awkward conflicts with demographic analysis that dogged the 1992 evaluation of DSE will go away. There will no longer be an independent DA estimate of undercount that can be compared to prior censuses because there won't be an actual enumeration figure to compare with. The long time-series of this single most trusted measure of undercount will be broken. But the questions about the accuracy of demographic analysis will disappear as well -- the subtle discrepancies that have allowed DA assumptions to be tested and refined over many decades will be lost in a sea of additional sampling error, imputation error, and bias from DSE.

DSE will look better as well. A few million people attributable to bias doesn't look so bad against a backdrop of twenty-odd million uncounted people. But the bias will still be there. Adding a large chunk of more predictable uncounted cases will make DSE look better but it won't reduce the kernel of tough uncounted cases that brought DSE to its knees in the 1990 Census. In fact, since major portions of the bias arose from DSE operations and procedures, they will scale right up with the artificially inflated total of the uncounted. And on top of this, the strategy for truncating follow-up will add sampling error and imputation error in millions of cases where full follow-up would have produced accurate data.

If Congress can't find the resources to intensively follow-up every citizen who can be convinced to participate in the census, it deserves the inaccurate census it will get. You will witness the ultimate triumph of calculated error over actual data.

I thank the Subcommittee for this opportunity to express my comments and would be pleased to respond to questions.