Understanding Factors that Contributed to the Large Error of Closure in Census 2000
by J. Gregory Robinson and Kirsten K. West, U.S. Bureau of the Census

The purpose of this paper is to put in perspective the relative contribution of the two components of the large 6.8 million error of closure, which resulted when the 1990 census-based population estimate of 274.6 million was compared to the Census 2000 count of 281.4 million :

* the amount attributable to change in coverage from 1990 to 2000
* the amount attributable to error in the 1990 census-based population estimates for 2000.

Many users have the perception that the error of closure was largely caused by the understatement of growth by the population estimates for the 1990's. Now that the coverage studies for 2000 have been completed, we can systematically assess the contribution of estimates error and coverage change. The analysis would analyze the error of closure by age, sex, and origin (Hispanic, Not Hispanic). Two sets of population estimates would be utilized: (1) the original estimates for 2000 prepared in advance of the census and (2) estimates re-calibrated with revised components of population change, based on the Demographic Analysis and Population Estimates (DAPE) work after the census.

In sum, studies are available that evaluate separately each component of the error of closure in Census 2000; in this paper the components are examined jointly to provide an empirically-based explanation for the shortfall of the population estimates compared to what the census counted.

Results:

About 60 percent or more of the 6.8 million error of closure is attributable to the dramatic reduction in net undercount from 1990 to 2000. Both the Demographic Analysis estimates and Accuracy and Coverage Evaluation results measure a reduction of 4 million or more in the net undercount. The remainder of the error of closure includes that attributable to understatement of growth by the population estimates.

More importantly, when examined by origin, the error of closure affected the estimates for Hispanics much more than the estimates for Non-Hispanics. The error of closure was 3.2 million, or 8.9 percent for the Hispanic population, but only 1.5 percent for Non-Hispanics. The reworked DAPE estimates more than halve the error of closure for Hispanics (from 3.2 to 1.5 million, or 4.2 percent), and the error for Non-Hispanics drops marginally to 1.1 percent. This analysis with the DAPE results is significant in that it identifies that the understatement of growth (largely the understatement of net international migration) is confined mostly to the Hispanic population estimates.

The relative contribution of estimation error to the overall error of closure can be carefully assessed only when the change in coverage is taken into account. For Non-Hispanics, the error of closure is disproportionately explained by the dramatic change in coverage from 1990 to 2000; understatement of growth of the Non-Hispanic population was not a major factor, with or without the DAPE estimates. For Hispanics, however, the understatement of growth during the 1990's and appreciable change in coverage were both factors contributing to the very large

closure error for Hispanics.

Conclusion:

The error of closure in Census 2000 is largely attributable to a large change in census coverage in 2000 relative to 1990. The understatement of population growth (and underlying net immigration estimates) during the 1990's primarily affected the population estimates for Hispanics; the population estimates for Non-Hispanics were reasonable.

Looking ahead to 2010, two lessons come from the analysis of the error of closure in 2000. First, we need to improve our estimates of population change for all groups, and the vigilant monitoring of change with the American Community Survey and other sources will help improve the estimates for 2010. Second, we should anticipate changes in coverage that may occur in the 2010 census and assess how this may impact the measured error of closure even before the census takes place. For instance, if the 6 million or so estimated duplicates in Census 2000 are nearly eliminated in 2010, this will automatically cause the population estimates to appear to be 6 million "too high" relative to the census. The estimates could in fact be quite accurate–the shortfall of the census relative to the estimates would be a consequence of coverage change (opposite the direction of the error of closure in 2000).