

## DISCUSSION: THE 2000 CENSUS ADJUSTMENT DECISION

Alan M. Zaslavsky

Department of Health Care Policy, Harvard Medical School, Boston, MA 02115

**Keywords:** census undercount, coverage measurement

### Background

Undercount in the U.S. Decennial Census, and the possibility of correcting for differential undercount through adjustment, has been controversial for at least 3 decades. Each of the last three censuses has made some progress toward estimating and correcting the undercount. In the 1980 Census, a systematic effort was made to estimate undercoverage, relying in part on data from the Current Population Survey. Although useful estimates were obtained, the Census Bureau's technical staff largely defended the position that they were not of adequate precision to be used for adjustment. The Post Enumeration Survey (PES) implemented as part of the 1990 census was designed to support adjustment, although in the political struggles preceding the census its sample size was halved from 300,000 to 150,000 households. Subsequently the Director of the Census Bureau recommended use of adjusted counts but was overruled by the Secretary of Commerce.

Research during the 1990-2000 intercensal decade developed a procedure for estimation of undercount based on an Accuracy and Coverage Evaluation survey (ACE) broadly similar to the 1990 PES, and a set of decision criteria, to be applied on a highly accelerated timetable after the census.

At the beginning of March, 2001, the Secretary of Commerce had to decide whether to release adjusted or unadjusted counts for use in redistricting by the states. (The Supreme Court had already forbidden the use of sampling for apportionment of representatives among the states, based on interpretation of ambiguous and contradictory language in the census legislation.)

Just before this deadline, the Census Bureau issued a recommendation against using the adjusted numbers, primarily due to disagreement on differential undercounts between the results of the ACE and those of demographic analysis (DA). This discrepancy was given great weight for two reasons: (1) DA was the only alternative quantitative standard against which the accuracy of the census and ACE results could be evaluated at that time, and (2) the fairly close agreement between estimates using the PES and those from DA had been regarded as among the most striking pieces of confirming evidence for the 1990 PES estimates. The Secretary's decision followed this recommendation. I was one of the informal group of experts asked to advise the Secretary on this decision. As such, I personally was in the awkward position of having advise against adjustment on the general principle that it was inappropriate to ask the Census Bureau to implement a process it did not feel it could stand behind, even though the indications

were that the ACE had been procedurally successful and the discrepancies with DA might have alternative explanations. Research to resolve the inconsistencies in the results has continued and as of these meetings was still underway.

In my comments on John Thompson's excellent and comprehensive presentation, I note some of the most striking results of the recent undercount estimation process, and offer some views on the appropriate standards to be applied in deciding whether to use statistical methods for adjustment.

### Interesting results

A number of striking results emerged from the 2000 ACE, especially when considered in relationship to historic trends in census undercount. First, estimated racial (black/white) differentials were sharply reduced from historical levels of about 4% typical of the past 6 decades (Passel, 1991; Hogan, 1993) to about 2%. The differential undercount for children was also sharply reduced. The similarity of ACE measurement methods to those used in the 1990 census lends credence to interpreting this as a real change. It may be attributed to improved census methodology, such as improved techniques for enumeration of children in enumerated households, to outreach and partnerships in minority communities, and perhaps to social changes that may have affected racial differentials in the undercount.

On the other hand, differences based on tenure (homeowners versus renters), identified as an important predictor of undercount in 1990, remained large. This difference might be a proxy for housing unit characteristics that interact with census processes to exacerbate undercount, such as multiunit dwellings. It also might be simply the best available proxy for social class differences that affect participation. Further analyses linking underenumeration to household and housing unit characteristics from the long form and Master Address File may help to explain this difference.

### Accuracy evaluation

The main quantitative check on the ACE results was the comparison to DA. To everyone's surprise, DA estimates fell substantially below those from both the unadjusted census and the census adjusted by ACE. This finding naturally focused attention on the accuracy of DA estimates, particularly estimates of immigration.

Passel (2001) argues that increases in undocumented immigration, together with declines in emigration associated with the boom economy of the late 1990s, substantially increased the number of immigrants beyond levels estimated under assumptions that had been plausible in previous censuses. The Census Bureau's decision process included a "high" DA

estimate based on doubling the assumed contribution to population due to undocumented immigration, but still fell short of ACE estimates. If doubling the estimate is a plausible alternative, though, might not tripling also be plausible? This would be more than enough to match the ACE's estimates of Hispanic population. The Census Bureau's demographers argued that such a modification of the immigration estimates would generate age distributions inconsistent with those observed. Such an argument, however, is sensitive to assumptions about the age distribution of immigrants, and this distribution, like the overall number, could have changed under changing economic circumstances.

Another social change with implications for DA is the shift in patterns of racial identification. Former Census Bureau Director Kenneth Prewitt argues that when the history of the 2000 census is written, its most significant feature will be the change in racial identification questions. The "check all that apply" approach adopted by the Office of Management and Budget and first implemented in this census reflects an increasingly fluid sense of racial/ethnic identity in American society more broadly. The impact of these changes might go beyond the limited numbers who elected to check multiple racial categories.

If the way people report racial identity truly is changing, this has greater implications for DA estimates of the differential undercount than for those from survey-based methodologies like the ACE. The latter only requires a reasonable degree of consistency between the census and a survey, conducted several months apart. (As long as there is adequate identifying information for matching, inconsistency in racial identification only affects estimation through cases that did not appear in the census, a relatively small part of the ACE sample.) DA estimates of differential undercount, on the other hand, rely on projection of racial categories given at birth over the lifetimes of the cohort, including death records and (for comparisons to census estimates) current racial self-identification. (Also see Passel 2001.)

The evidence suggests that the ACE was operationally successful. Concerns about bias of ACE estimates center around the processes of ACE measurement, specifically balancing error and E-sample error. The magnitudes of their contributions to bias should be better known after evaluation studies have been completed; results of such studies could not have been ready in time for the March, 2001 decision.

One might ask to what extent the decision on adjustment depends on "Census esoterica", issues like balancing and E-sample error that can only be understood fully by a handful of professionals at the Census Bureau or outsiders willing to devote substantial efforts to absorbing specialized knowledge. On the other hand, although DA also involves technical issues, the most important issues for DA are not particularly arcane.

The magnitude of immigration and the nature of racial/ethnic self-identification are among the most important aspects of current social changes in our nation. Furthermore, alternative hypotheses on how these are changing will be generated by research conducted outside as well as within the Census Bureau. Hence, this topic should be the subject of broad scientific and public policy discussion. To facilitate this discussion, data that underlie the DA analysis should be made available for analysis. High quality, accessible documentation is also essential to make possible reproduction of the Bureau's analyses and tests of sensitivity to alternative assumptions. In the long run, special studies or even new measurement systems might be required to determine in which areas DA is accurate, and therefore preserve its usefulness. For example, it is generally agreed that DA is more accurate for predicting sex ratios than for absolute numbers, since some of the mechanisms causing error in DA affect males and females equally. This might be verified through longitudinal analyses of microdata samples, and similar analyses could evaluate error due to shifting racial identification. The Census Bureau currently is improving estimates of undocumented migration by analyzing country-of-origin responses from the Census 2000 Supplementary Survey. Discussion of these issues will improve our understanding of the composition and dynamics of the U.S. population as well as the accuracy of census counts.

#### **Decision issues**

Because the evidence on using the ACE to adjust small-area census counts is might never be conclusive, it is worthwhile to consider carefully the criteria for this decision. I regard some standards for evaluation as *inappropriate*. *Gross* error (omissions plus erroneous inclusions), rather than *net* error, has been suggested as a measure of accuracy of the census. The interpretation of gross error depends on how it is calculated under ACE procedures. Some erroneous enumerations and omissions represent properly enumerated cases that appear to be errors because of the criteria for matching in the ACE. For example, an error that placed a case in one block for the census and in a nearby block for the ACE could be treated as two gross errors, although with slightly different processing it might be detected as a correct enumeration. Thus, gross error for very small areas such as census blocks is not readily interpretable. On the other hand, a similar measure applied to large geographical or demographic units (as a sum of absolute errors) is a sensible summary of census accuracy.

Another inappropriate standard is correction of counts at the lowest possible level, the individual block. Even the best PES could not tell exactly how many individuals were missed or overcounted in each block – that would require a replication of the entire census. Fortunately, all important uses of the census involve

aggregates of blocks, even if block-level data must be released to facilitate flexible aggregation (as in legislative redistricting). Someday we might estimate census error down to the block level using improved administrative record systems, but our current inability to do so is not an obstacle to adjustment.

The latter problem is the *reductio ad absurdum* of the critique of synthetic estimation. Rather than asking whether the synthetic assumption holds, down to fractions of undercounted individuals, let us be guided by George Box's well-known admonition that "all models are wrong, but some are useful." The Census Bureau has properly chosen to focus on error for important units such as states or congressional districts. Reducing error for larger units typically suggests that error will be reduced for smaller units as well (Schirm and Preston 1987), although hypothetical counterexamples can be constructed. However, reduction of error for smaller units is more difficult to demonstrate due to the sparseness of the ACE sample and corresponding noisiness of the evaluation at that level. "Synthetic error" affects both the ACE and the unadjusted census, since the assumption of the synthetic model (uniform adjustment across each poststratum) is weaker than that of the "no-adjustment model" (uniform zero adjustment). Hence I would want to scrutinize carefully the contribution of "synthetic error" to the loss function comparisons and its implications for adjustment, focusing on accuracy of shares and on weighted measures of loss.

The decision process followed (almost) prespecified procedures, as intended. The acceleration of the decision schedule relative to that followed in 1990, driven by legislative deadlines for release of redistricting data, allowed little time for evaluation. Consideration should be given to relaxing these deadlines, should a future Congress be sympathetic to these needs of the Bureau.

The Secretary of Commerce's originally proposed criterion for the 1990 adjustment decision was essentially that the adjustment had to demonstrably improve accuracy at all levels of aggregation, an impossible criterion to meet for reasons outlined above. For 2000, the criteria were nominally that ACE results could be used if it were operationally successful, consistent with historical undercount patterns and DA population estimates, and supported by reviews of quality measures. In practice, the decision faced by the Census Bureau could be stated as: "Are you willing to bet the farm on this adjustment?" Thus, the Census Bureau had to take a conservative approach to the adjustment; as noted above, outside experts were then forced to concur.

Because ACE estimated a smaller differential undercount by race than in 1990 (although still a substantial differential by tenure), loss function analyses found it harder to demonstrate improvement due to adjustment; nonetheless, the evidence for improvement was strong. The main difference from 1990, however, was that the

discrepancy between ACE and DA removed the main external confirmation of the former, indeed questioning whether their agreement in 1990 was a coincidence.

The standard dichotomous "decision" framework is a limited paradigm driven by requirements of prespecification and simplicity. It would be preferable to combine all sources of information (including evaluations) to construct a "best" estimate (Zaslavsky 1993), either for production use, or to evaluate available alternatives.

Ultimately, the "adjustment decision" is political, involving competing principles of stability and equity, also reflected in the dissenting and concurring opinions in the 1999 Supreme Court decision. Justice Scalia argued that having a noncontroversial (=prespecified) apportionment system was the Founders' paramount concern, while Justice Stevens emphasized their concern with equity (=accuracy). The Bureau attempted to reconcile these values through a statistically sophisticated but prespecified procedure, but the political process might not be ready for such innovations.

At this writing, we await the Census Bureau's recommendation on use of adjusted estimates for purposes other than redistricting. At the very least, the Bureau should release the full results for discussion and research as soon as possible. Nonetheless, we must ask how the Census Bureau can ever take the risks involved in making more radical changes in the census process, such as more extensive use of administrative records.

In conclusion, I congratulate the Census Bureau on its magnificent implementation of the Census and ACE, on its openness, and on its dedication to seeking out sources of error and improving accuracy of the census.

#### Postscript

As this discussion was prepared for final submission on October 17, 2001, the Census Bureau announced that further research had found a large amount of previously undetected duplication in the census, leading to erroneously high estimates of population, and that population estimates would *not* be adjusted.

#### References

- Hogan, H. (1993), The 1990 Post-Enumeration Survey: Operations and results, *JASA*, 88, 1047-1060
- Passel, J.S. (1991), Age-period-cohort analysis of Census undercount rates for race-sex groups, 1940-1980: Implications for the method of demographic analysis, *ASA Proc. Social Statistics Section*, 326-331
- Passel, J.S. (2001), Demographic analysis, an Evaluation, in Census Monitoring Board, Presidential, *Final Report to Congress*, <http://www.cmbp.gov>
- Schirm, A.L., and Preston, S.H. (1987), Census undercount adjustment and the quality of geographic population distributions, *JASA*, 82, 965-978
- Zaslavsky, A.M. (1993), Combining census, dual-system, and evaluation study data to estimate population shares, *JASA*, 88, 1092-1105