

I am joined in this testimony by my colleague David Freedman. We think there is widespread—although by no means universal—agreement on two chief points. First, Census 2000 succeeded in reducing differential undercounts from their 1990 levels. Second, there are serious questions about the accuracy of proposed statistical adjustments. The Bureau advised the Secretary to certify the unadjusted counts and we concurred, as did the Secretary's other advisors.

Statistical adjustment faced a new problem in Census 2000. Independent population estimates are derived by Demographic Analysis from administrative records, including birth and death certificates and Medicare files.¹ These estimates show the Census overcounted the population by perhaps 2 million people. Proposed statistical adjustments would have added another 3 million people, making the overcounts even worse. Demographic Analysis is independent of the Census and the ACE survey which underlies proposed adjustments. Demographic Analysis and ACE point in opposite directions. While Demographic Analysis is hardly perfect, it is a stretch to blame Demographic Analysis for the whole of the discrepancy with ACE.²

Mistakes in statistical adjustments to the Census are nothing new. Our studies of the 1980 and 1990 data have described three kinds of error: processing error, correlation bias, and heterogeneity.³ In the face of these errors, it is hard for adjustments to improve on the accuracy of Census numbers for states, counties, legislative districts, and smaller areas. Statistical adjustments could easily put in more error than they take out, because the Census is already very accurate.

What went wrong with ACE in 2000? Errors in responses to the survey or in the statistical operations may from some perspectives have been under better control than they were in 1990. But, it appears, processing errors must have been worse in other respects. Research is underway to pinpoint the difficulty. The Bureau is investigating a form of error called balancing error.⁴ We suspect that troubles also occurred in a new treatment of movers and in the detection of duplicates, which were especially numerous in 2000.

Correlation bias and heterogeneity are endemic problems that make it extremely difficult for adjustment to improve on the Census. Correlation bias is the tendency for people missed in the Census to be missed by ACE as well. Correlation bias in 2000 may have amounted, as it did in 1990, to millions of persons. These people cannot be evenly distributed across the country. If so, statistical adjustments create a distorted picture of census undercounts.⁵

Heterogeneity means that undercount rates differ from place to place within population groups treated as homogeneous by adjustment. Heterogeneity puts limits on the accuracy of adjustments for areas like states, counties, or legislative districts. Our studies, along with recent work at the Bureau, show that heterogeneity remains a serious concern.⁶

Census 2000 achieved a high level of accuracy. Given that, and given the problems with statistical adjustments, the Secretary's decision to certify the census counts was the right decision.

Notes

1. The Bureau's estimates from Demographic Analysis are presented in the B-4 Report; see especially Appendix Table 2. The estimated total national populations are as follows:

Demographic Analysis	279.598 million
Census 2000	281.422 million
ACE	284.684 million

2. The Demographic Analysis estimates for net undocumented immigrants and some other categories of non-citizen residents may be somewhat low. But, as the B-4 Report spells out on page 11, implausibly large revisions would be required to bring the totals into agreement with ACE.

3. See, e.g., Freedman and Navidi (1992), Freedman and Wachter (1994), and Brown et al. (1999).

4. The Bureau's research on balancing error is summarized on pages 24–25 of the ESCAP report.

In July 1991, the Bureau recommended adjusting Census 1990 by adding 5.3 million people: processing errors were estimated at 1.7 million (these figures are net, nationwide). The figure for processing error later increased to 3.0 million, although independent estimates range up to 4.2 million. Our estimate is 3.6 million (Wachter and Freedman, 2000a).

In 2000, ACE would add 3.3 million persons to the census count. The preliminary estimate of processing error is 2 million, as may be seen by doing some arithmetic on the percentages in Table 24 of the B-19 Report. That table allows 1 million for correlation bias. However, the underlying model (B-12) repeats the error discussed in Wachter and Freedman (2000a). Compare page 16 of B-12, although page 46 of the B-1 Report acknowledges the problem.

Thus, the estimate for correlation bias needs to be increased, perhaps by another million or so. If the population estimates from Demographic Analysis are approximately correct, the estimates for processing error will need to increase by several million, as in 1990.

5. Brown et al. (1999), Wachter and Freedman (2000a).

6. Adjustment assumes that coverage rates (i.e., rates of census undercount or overcount) are constant within population groups called "post strata," across wide stretches of geography. Failures in this assumption are called heterogeneity, or called "synthetic error" in the B-studies. The Bureau's research on heterogeneity is summarized on pages 22–24 of the ESCAP report. Our work is described in Freedman and Wachter (1994) as well as Wachter and Freedman (2000b).

References

Anderson, M. and S. E. Fienberg (1999). *Who Counts? The Politics of Census-Taking in Contemporary America*. Russell Sage Foundation, New York.

Breiman, L. (1994). The 1991 census adjustment: undercount or bad data? *Statistical Science* 9: 458–537 (with discussion).

Brown, L. D., M. L. Eaton, D. A. Freedman, S. P. Klein, R. A. Olshen, K. W. Wachter, M. T. Wells, and D. Ylvisaker (1999). Statistical controversies in Census 2000. *Jurimetrics* 39: 347–75.

- Cohen, M. L., A. A. White, and K. F. Rust, eds. (1999). *Measuring a Changing Nation: Modern Methods for the 2000 Census*. National Academy Press, Washington, D. C.
- Fay, R. E. and J. H. Thompson (1993). The 1990 Post Enumeration Survey: statistical lessons, in hindsight. *Proceedings, Bureau of the Census Annual Research Conference*. Bureau of the Census, Washington, D. C.
- Freedman, D. A. and W. C. Navidi (1992). Should we have adjusted the U. S. census of 1980? *Survey Methodology* 18: 3–74 (with discussion).
- Freedman, D. A., P. B. Stark, and K. W. Wachter (2000). A probability model for census adjustment. <http://www.stat.berkeley.edu/~census/557.pdf> To appear in *Mathematical Population Studies*.
- Freedman, D. A. and K. W. Wachter (1994). Heterogeneity and census adjustment for the intercensal base. *Statistical Science* 9: 476–485 (with discussion).
- Freedman, D. A., K. W. Wachter, D. R. Cutler, and S. P. Klein (1994). Adjusting the census of 1990: loss functions. *Evaluation Review* 18: 243–280.
- Prewitt, K. (2000). Accuracy and Coverage Evaluation: statement on the feasibility of using statistical methods to improve the accuracy of Census 2000. *Federal Register* 65: 38373–38398.
- Skerry, P. (2000). *Counting on the Census*. Brookings, Washington, D. C.
- U. S. Census Bureau (2001). Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy. Cited as the ESCAP report. The supporting documentation consist of reports B-1 through B-19. All these reports are available on the web. <http://www.census.gov>
- Wachter, K. W. and D. A. Freedman (2000a). The fifth cell. *Evaluation Review* 24: 191–211.
- Wachter, K. W. and D. A. Freedman (2000b). Measuring local heterogeneity with 1990 U. S. Census data. *Demographic Research*, an on-line journal of the Max Planck Institute. Volume 3, Article 10. <http://www.demographic-research.org/Volumes/Vol3/10/>