Why Is There Still a Controversy About Adjusting the Census for Undercount?

Margo Anderson, *University of Wisconsin, Milwaukee*
Stephen E. Fienberg, *Carnegie Mellon University*

We appear to agree with Brunell (2001) on four factual matters:

1. "Census 2000 is over; the debate over adjusting the census is not" (Brunell 2001).
2. The computer coding error in 1990 would not have erroneously shifted any congressional seats has adjusted data been used for reapportionment.
3. In March 2001, the Census Bureau announced that it was "unable to conclude . . . that the adjusted Census 2000 data are more accurate for redistricting."
4. The Census Bureau has refused to release the adjusted 2000 census data.

Otherwise we find Brunell's arguments against census adjustment either distorted or confused.

With regard to the 1990 coding error, he persists in confounding it with other Census Bureau revisions to the adjusted data. Even one of his favorite sources of adjustment criticism, Brown et al. (1999), has become more cautious!

In both 1990 and 2000, the Census Bureau took a sample survey in conjunction with the April enumeration, expecting to issue revised population estimates for local areas based on the results of the survey. In 1990 the survey was called the Post Enumeration Survey (PES); in 2000 it was the Accuracy and Coverage Evaluation survey (ACE). The bureau intended to use the survey to reduce the known level of error in the census.

In both 1990 and 2000, this effort was surrounded by technical and political controversy (Anderson and Fienberg 2001b). In 1990, the census director recommended that the adjusted census results be certified as the official results. The secretary of commerce rejected the recommendation and certified the uncorrected data from the April enumeration as the official census results. In 2000, the acting director of the census bureau recommended that the uncorrected data from the April enumeration be certified as the official census results. In both censuses, the bureau compared the adjusted data with the unadjusted counts using a "total error" model and loss function analyses.

Despite Brunell's claim to the contrary, these loss function analyses included the bureau's best estimates of heterogeneity bias (this is the correction for correlation bias in the quote he provides!). In 1990 none of the bureau's analyses showed that the unadjusted data were superior at any level of geography. In 2000, only one out a large number of analyses suggested the superiority of the raw census counts, and this was only for that subset of counties with population under 100,000, identified post hoc. The analyses addressed at congressional reapportionment all suggested the superiority of the adjusted data. Brunell refers the reader to two reviews of our book and to Breiman (1994)—which do not address these analyses or present alternatives in any form—and to Brown et al. (1999), whose discussion of loss function analyses is ridded with unjustifiable assumptions and errors (see Anderson et al. 2000).

Brunell (2001) raises a new issue of "synthetic error," which relates to the application of adjustment factors computed from the ACE sample to the nation as a whole. While this was not included in the total error model and loss function analyses in either 1990 or 2000, the bureau carried out separate studies of its likely impact. It concluded that synthetic error was unlikely to have any serious effect on the analyses. So our answer to Brunell is that we take ACE error seriously, as does the Census Bureau. Both the census and the adjusted census data have their imperfections. The question is: Which appears to be doing a better job?

Are the results we cite definitive? Clearly not, since the recently released October 2001 ESCAP II (U.S. Bureau of the Census 2001a) reports raise major new questions about the accuracy of local area Census 2000 counts. But the ACE-adjusted counts are also not flawed by error and assumption in the way Brunell suggests.

Why is there so much confusion? In both 1990 and 2000, during the controversies surrounding the proposals to adjust the census, the Census Bureau generated a variety of census results and estimates of census error. In the welter of discussion about adjustment, it is sometimes difficult to determine which data and estimates are being discussed, and how they should be assessed. It is perhaps not surprising that Brunell is confused.

Table 1 provides estimates of the 1990 and 2000 population and two measures of census error, net error, and gross error. It also provides the estimates of the 1990 and 2000 population and error from two methods used to measure error, i.e., demographic analysis and dual-systems estimation using the PES or the ACE.

Editor's Note
This is the last in a series of commentaries on the 2000 Census between the authors here and Thomas Brunell (article following). This discussion first began in the December 2000 issue of *PS*, with articles that explored the Census and the political issues surrounding its interpretation. For a look at the complete set of article please visit *PSonline* at www.apsanet.org/PS/dec00/.

https://doi.org/10.1017/S1049096502000185 Published online by Cambridge University Press

Table 1
Estimates of Census Error in the U.S. Population in Millions, 1990 and 2000

Estimate	Population	Net Error N	Net Error %	Gross Error N	Gross Error %	Release Date
1990 Census						
1990 Official	248.7					Dec 1990
1990 DA	253.4	4.7	1.9	Does not measure		1991
1990 Revised DA	252.9	4.2	1.65	Does not measure		1993
1990 PES	254.0	5.3	2.1	25.7	~10	July 1991
1990 PES	252.9	4.2	1.7	Not available		Jan1992
1990 Revised PES	252.7	4.0	1.61	12.8–26.4	~5–10	1993
2000 Census						
2000 Official	281.4					Dec 2000
2000 DA	279.6	–1.8	–.65	Does not measure		March 2001
2000 Alternate DA	282.3	.9	.32	Does not measure		March 2001
2000 Revised DA	281.8	.34	.12	Does not measure		Oct 2001
2000 ACE	284.7	3.3	1.18	9.5–15.1	3.4–5.4	March 2001
2000 ACE Revised	281.6	.169	.06	Not available		Oct 2001

Sources: Hogan 1993; Robinson et al. 1993; U.S. Bureau of the Census 2001a, 2001b. The comparison of the July 1991 and January 1992 adjusted census counts for states are available in Defendant's Exhibit 21, Memorandum for Charles D. Jones from John R. Thompson, February 5, 1992, New York v. Department of Commerce.

Net error is defined as the difference between the omissions in the census and the erroneous enumerations in the census. Leaving aside some complications, erroneous enumerations include duplicate records, fictitious records, and people counted in the wrong place or at the wrong residence (Anderson and Fienberg 2001a). Gross error is defined as the sum of omissions and erroneous enumerations.

The official 1990 population was 248.7 million. The Census Bureau reported that figure was an underestimate of between 4.0 and 5.3 million; the final bureau estimate of net undercount was 4.2 million using demographic analysis and 4.0 million using PES. In 2000, the official population count is 281.4 million. Estimates of census error released thus far range from an undercount of 3.3 million to an overcount of 1.8 million, and the wide range of estimates has cast a cloud on both Demographic Analysis (Citro, Cork, and Norwood 2001 and ACE as methods of measuring census error. The current bureau estimates are an undercount of 340,000 from demographic analysis and 169,000 from ACE (if we correct it for additional erroneous enumerations without altering the census counts at the same time).

The estimates of gross error in both 1990 and 2000 are much higher, in the range of 12.8 to 26.4 million in 1990 and at least 9.5 to 15.1 million for 2000 (the low estimates are the ones Brunell cites and they purposely exclude geographic errors). Estimates of gross error are only available from PES or ACE. Though bureau officials and observers generally have understood for years that erroneous enumerations and omissions might not balance out in local areas, there was no refined way of measuring gross error until the advent of PES and dual-systems estimation methods of the late 1980s. In other words, the 1990 Census was the first census for which reasonable measures of gross error were available. The bureau research tradition has focused on reducing net error, and how it was distributed differentially to minority groups such as African Americans and Hispanics, not on gross error.

Estimates of gross error derived from the ESCAP I analyses released in March suggest that gross error was reduced between 1990 and 2000. But in October, Acting Census Director William Barron reported that preliminary results of new tests had identified an additional 3 million erroneous enumerations in the census. If these estimates are confirmed in future research, the level of gross error will approach that from 1990 and the results will challenge many of the assumptions about the measurement of error in the census. For example, net error is clearly an inappropriate measure of error if the census contains a higher level of erroneous enumerations than previously known. Does gross error show the same patterns of geographic and minority concentration as net error was assumed to have in previous censuses? And what impact would such a rethinking of the measure of census error have on the understanding of the uses of the census for funding allocation, redistricting, etc.? And finally, will the PES/ACE methodology have to be rethought if it is not capturing the proper level of erroneous enumeration in the census? Answering these questions will require research.

As Brunell (2001) noted, the Bureau has refused to release local area ACE adjusted data. Since November 2001, two federal district courts have found for plaintiffs asking for the release of the data.

References

Anderson, Margo and Stephen E. Fienberg. 2001a. "Counting and Estimation: Methodology for Improving the Quality of Censuses: The U.S. 2000 Census Adjustment Decision." Presented at the International Conference on Quality in Official Statistics, Stockholm, Sweden. <http://lib.stat.cmu.edu/~fienberg/CensusAdjustment-Q2001.pdf>.

https://doi.org/10.1017/S1049096502000185 Published online by Cambridge University Press

——. 2001b. *Who Counts? The Politics of Census Taking in Contemporary America.* Revised paperback edition. New York: Russell Sage Foundation.
Anderson, Margo, Beth Daponte, Stephen E. Fienberg, Joseph B. Kadane, Bruce D. Spencer, and Duane Steffey. 2000. "Sample-Based Adjustment of the 2000 Census—A Balanced Perspective." *Jurimetrics* 40:341–56.
Breiman, Leo. 1994. "The 1991 Census Adjustment: Undercount or Bad Data?" *Statistical Science* 9:458–75.
Brown, Lawrence D., Morris L. Eaton, David A. Freedman, Stephen P. Klein, Richard A. Olshen, Kenneth W. Wachter, Martin T. Wells, and Donald Ylvisaker. 1999. "Statistical Controversies in Census 2000." *Jurimetrics* 39:347–75.
Brunell, Thomas. 2001. "Census 2000—Epilogue." *PS: Political Science & Politics* 34.
Citro, Constance F., Daniel L. Cork, and Janet L Norwood, eds. 2001. *The 2000 Census: Interim Assessment.* Panel to Review the 2000 Census, Committee on National Statistics, National Research Council. Washington, DC: National Academy Press.
Hogan, Howard. 1993. "The 1990 Post-Enumeration Survey: Operations and Results." *Journal of the American Statistical Association* 88:1047–60.
Robinson, J. Gregory, Bashir Ahmed, Prithwis DasGupta, and Karen A. Woodrow. "Estimation of Population Coverage in the 1990 United States Census Based on Demographic Analysis (with discussion)." *Journal of the American Statistical Association* 88:1061–79.
U.S. Bureau of the Census. 2001a. "Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy on Adjustment For Non-Redistricting Uses" (October 17). Washington, DC: Department of Commerce.
——. 2001b. "Report: Recommendation Concerning the Methodology to be Used in Producing Tabulations of Population Reported to States and Localities Pursuant to 13 U.S.C. 141(c)" (March 1). Washington, DC: Department of Commerce.

Why There Is Still a Controversy About Adjusting the Census

Thomas L. Brunell, *Binghamton University*

I will be brief. There is still a controversy over adjusting the census because the statistical adjustment of the census did not work in 1990, nor did it work in 2000. The adjustment process, in both theory and practice, is complicated and flawed. The chance of it working in future censuses without major changes is small. Anderson and Fienberg continue to twist language and report partial results in order to confuse the debate. For instance, even though, as Anderson and Fienberg argue, the computer coding error by itself may not have shifted a congressional seat in 1990, the computer coding error *along with the other errors* associated with the 1990 adjustment *would have shifted a seat.* They ignore the other errors associated with the 1990 PES. Considering the computer error in isolation is meaningless. Furthermore, they also gloss over that the computer coding error would have added roughly one million nonexistent people to the census—this is a problem regardless of the distribution of congressional seats.

Relying on the statistical adjustment to fix all the ills of the census is not wise. The feeling I got from Census Bureau officials when I was on Capitol Hill was that they thought the statistical adjustment would clean up all of the errors associated with the headcount. Thus, the predominant attitude was that it was not critical for the headcount to be as accurate as possible since the adjustment would sweep through, cleaning up all errors. The headcount, and not the adjustment, should be the focal point for census accuracy. Ironically, a census with an adjustment ultimately leads to a less accurate headcount simply because the post-census adjustment becomes a crutch. Combine that with the adjustment's adding as many errors as it takes out and it becomes fairly obvious why the adjustment process is flawed.

Worse still, there is not a single standard statistical adjustment model. As documented earlier in this debate, the Census Bureau faces dozens (Hundreds? Thousands?) of decisions about the nature of the statistical model used to make population adjustments. Each of these decisions can affect the results. If Census 2010 incorporates a statistical adjustment, I predict there will be a battle royal in Congress over what the adjustment ought to look like. Look for congressional hearings with hot topics such as: why the regression-smoothing model is more appropriate that the method that uses iterative proportion fitting; or why the statistical model with 356 post-strata is more appropriate than the model with 712; etc. Questions with political implications will be answered politically.

How then do we make the census better? I think the accuracy of the census ultimately depends on the involvement of local constituencies. Rather than a centralized census "trickling down" from Washington to the rest of the country, the process ought to "bubble up" from bottom to top. Localities have to understand the importance of the census and do what they can to increase the accuracy of the count. The federal government and the Census Bureau need to maximize the involvement of local constituencies in census taking, rather than responding to complaints about the accuracy of the results after the census is over.

https://doi.org/10.1017/S1049096502000185 Published online by Cambridge University Press