Census 2000—Epilogue*

Thomas L. Brunell, *Binghamton University*

Census 2000 is over; the debate over adjusting the census is not. In this essay I respond to a number of points from Anderson and Fienberg (2001) and report some of the results from Census 2000.

First, in terms of my (Brunell 2000a, 2000b, 2000c) disagreements with Anderson and Fienberg (2001), one of the major points of contention has to do with the "computer coding error," which—as Anderson and Fienberg agree—would have erroneously added over one million people to an adjusted census in 1990. Would this error, considered in isolation, have switched a congressional seat from Pennsylvania to Arizona? Anderson and Fienberg say "no," but this seems like a smokescreen. After fixing the coding error, the Census Bureau also corrected the data for the 104 most extreme blocks in the Post Enumeration Survey (or PES, the 1990 adjustment procedure); tossed out the smoothing model it was using; and revised the post-stratification. These changes—fixing the coding error is by far the biggest—indeed switched the seat.

Next, Anderson and Fienberg (2001) criticize my portrait of the capture/recapture statistical model, mainly because I oversimplify the process. In my original article (Brunell 2000a), I used broad brushstrokes to paint a picture of how the adjustment works. The model in Billard (2000) adds a layer of complexity, but her sketch is still just that — a sketch. I left out much of the complexity of the model, as the "real" Dual System Estimator (DSE) model — called the Accuracy and Coverage Evaluation (ACE) in the 2000 Census — is far more complicated. For the reader truly interested in the statistical minutiae, I recommend the Census Bureau's "Detailed Specifications for A.C.E. Methodology" (5 volumes, 971 pages), which provides a comprehensive account of the inner workings of the model, both theoretically and operationally.[1]

In an earlier essay (Brunell 2000b), I emphasize that it is unwise to expect a statistical model to correctly adjust population totals for small geographic units. In my argument I refer to the ecological fallacy in the census adjustment. Professors Anderson and Fienberg object to my use of the ecological fallacy in this debate because adjustments are not made at the individual level, but rather at the block level. Technically, it would be better to refer to this as a "synthetic fallacy," rather than an ecological one. The statistical adjustment assumes error rates in the census are constant within post strata. This assumption is, of course, wrong. For instance, undercount rates vary within post strata by region. Asians living in Laguna Beach, California, will be missed less often than Asians living in Queens, New York. Population adjustments using the DSE make no distinction between these two groups; this is the synthetic error, which Anderson and Fienberg gloss over.

Anderson and Fienberg state that the adjusted data proved superior to the unadjusted data at the congressional district level in both 1990 and 2000. In Navarro and Asiala's (2001) study the news is quite mixed for increases in accuracy, and is worth quoting at length.

> "For levels the re/sults are mixed. The census is more accurate for states and counties for small levels of correlation bias. As the degree of correlation bias in the DSE increases—50% and higher—the ACE becomes more accurate than the census for states and countries. The census is more accurate than the ACE for counties with less than 100,000 population" (Navarro and Asiala 2001, 1).

It is clear from this quotation that the statistical models used to evaluate accuracy are based on assumptions (i.e., the level of correlation bias), and there is no proof that the adjusted data are more accurate. Anderson and Fienberg (2001) conclude, "In 1990, and again in 2000, adjusted data proved to be superior to the unadjusted counts." This is wrong. Not only do the analyses not support those kinds of conclusions, but the analyses themselves are suspect. Freedman and Wachter (2001, 6) argue that these analyses tend to favor the adjustment because, "Loss function analysis turns out to depend more on wishful assumptions than on data." For instance, the loss function analysis assumes no heterogeneity bias, which was discussed above as one of the central problems with the statistical adjustment. Navarro and Asiala (2001, 1) indicate: "The census error is measured by the net undercount. The census coverage is also subject to heterogeneity across areas. *This error is not quantified and is not included in the Total Error analysis*" (my emphasis).

Brown et al. (1999) figure that, "Even at the state level, the case for adjusting the 1990 Census was shaky at best: there was a strong likelihood that adjustment would have put in more error than it took out, and conclusions depend heavily on assumptions." It is rash to conclude that the adjusted data are superior to the unadjusted data, especially since the scientific evidence supports the opposite conclusion.

Anderson and Fienberg also state: "No statistical analyses showed that the [unadjusted] counts were superior to the adjusted counts at any level of geography." This too is wrong. I recommend reading Breiman (1994) and Brown et al. (1999). Most of the "fix" in the 1990 PES actually added new errors into the census rather than correcting existing ones. The method makes some corrections to the headcount, but it also inserts new errors. This is not a myth, and it remains the central problem for adjusting the census.

Objections to adjusting the census are well founded. To say that the adjustment changes the results of the census is obviously true. What is not obvious, but no less true, is that the adjustment procedure is just as likely to create additional errors in the census as it is to fix existing ones.

Results of Census 2000

Given that the 2000 Census is now complete, it is appropriate to review the results. The reader may know that the unadjusted data were certified as the official redistricting data for the states to use. Some people might assume, since George W. Bush ascended the Presidency in January of this year, that the new Republican administration put a stop to the adjustment. This is not true. The professional staff at the Census Bureau made the decision that the adjusted data were not accurate and would therefore not be released.

On March 1, 2001, William G. Barron, acting Director of the Census, issued a memo (Barron 2001) to the Secretary of Commerce, stating that the ESCAP[2] had concluded that the adjusted data for the census were not reliable enough to use for redistricting (ESCAP 2001). Prior to this announcement,

https://doi.org/10.1017/S1049096501000749 Published online by Cambridge University Press

observers assumed that the Census Bureau would likely release the adjusted data and then the Secretary of Commerce might have overruled this decision. However, the Bureau itself came to the conclusion that there were enough problems with the adjusted data that releasing them would be a mistake. The question then became "What went wrong with the adjustment in 2000?" It is too early to tell exactly what happened, but it was clear something was wrong when former director Prewitt announced that the raw headcount was over 281.4 million people—far higher than informed observers expected it to be *ex ante*.

There are two methods for checking the results of the census: the Dual System Estimator (DSE); and demographic analysis (DA), which uses birth, death, immigration, and emigration records to approximate the population of the country. Prior to the announced results of the census the expectations were roughly the following: DA estimate for the population was 279.6 million people. The census, was expected to come in lower than that, because it misses more people than it counts twice. Then the DSE would "clean up" the census by adding three to four million people and taking out one million people or so. The raw head count turned out to be 1.8 million *higher* than the DA estimate. This means that the census had a net overcount, rather than the expected net undercount. The DSE would have added a net 3.3 million more people to the census. "Fixing" the census with the results of the DSE would have then exacerbated the overcount. These results left the Census Bureau in a rather difficult position. On what basis would Bureau officials defend the adjustment? It is difficult to rely solely on the results of the ACE to defend the adjustment, since the ACE serves both as an indication of errors in the census, and the basis by which these errors are "fixed."

Is the Adjustment Dead?

The statistical adjustment of the 2000 Census is not dead, although there is no guarantee that the data will be used for anything in terms of real public policy. In fact, the Census Bureau shows little interest in releasing data that still contain significant problems that it does not fully understand. John H. Thompson, who directed the 2000 Census and was a major proponent of the adjustment, said: "We don't really want to release data that we don't have confidence in."[3] Democrats in Congress are keen on releasing the adjusted data into the public sphere so that the data may be used for redistricting. This would undoubtedly inspire lawsuits, but litigation in the redistricting process is the rule (without exception). There will be hundreds of studies focusing on problems with the statistical adjustment of Census 2000; perhaps they will shed some light on methods for improvement.

Anderson and Fienberg are less interested in taking on the problems associated with the adjustment (processing errors, heterogeneity bias, etc.), and more interested in focusing on errors associated with the head count. My question to them is: What went wrong in Census 2000? I am willing to agree that there are errors in the census, although the figure cited by Anderson and Fienberg of 25 million errors in the 1990 Census is grossly inflated. The official estimates from the Census Bureau indicate a "Gross Coverage Error" of 12.8 million people in 1990 and less than 10 million people in 2000 (U.S. Department of Commerce, Bureau of the Census 2001). Anderson and Fienberg are, again, propagating a few myths of their own (for additional examples of this see Stark 2001 and Ylvisaker 2001). The important question for Anderson and Fienberg to address is: What about the errors in the adjustment?

Notes

* I would like to thank David Freedman, Mike Miguel, Dave Clark and Tom Muller for their comments. Any remaining errors are my own.

1. Available online at <www.census.gov/dmd/www/DetailSpec.htm>.

2. ESCAP is the Executive Steering Committee for Accuracy and Coverage Evaluation Policy, an internal Census Bureau committee that evaluated the adjustment process. Their report opens with the following sentence: "The Executive Steering Committee for A.C.E. Policy (ESCAP) is unable to conclude, based on the information available at this time, that the adjusted Census 2000 data are more accurate for redistricting. Accordingly, ESCAP recommends that the unadjusted census data be released as the Census Bureau's official redistricting data."

3. Eric Schmitt, "Critics Want Census Data Made Public," *New York Times*, 26 June 2001, College Times.

References

Anderson, Margo, and Stephen E. Fienberg. 2001. "Getting the Facts on the Census Correct: A Response to Brunell." *PS: Political Science & Politics* 34:651–52.

Barron, William G. 2001. Recommendation on Adjustment of Census Counts. Memorandum, U.S. Bureau of the Census.

Billard, Lynne. 2000. "The Census Count: Who Counts? How Do We Count? When Do We Count?" PS: Political Science and Politics 22: 767–74.

Breiman, Leo. 1994. "The 1991 Census Adjustment: Undercount or Bad Data?" *Statistical Science* 9:458–537.

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 L. 2000a. "Using Statistical Sampling to Estimate the U.S. Population: The Methodological and Political Debate Over Census 2000." *PS: Political Science & Politics* 33:775–82.

——. 2000b. "Rejoinder to Anderson and Fienberg." *PS: Political Science & Politics*. 33:793–94

——. 2000c. "Making Sense of the Census: It's Political." *PS: Political Science & Politics*. 33:801–2.

ESCAP. 2001. "Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy: Recommendations Concerning the Methodology to be Used in Producing the Tabulations of Population Reported to States and Localities Pursuant to 13 U.S.C. 141(c)." Washington, DC: Department of Commerce.

Freedman, David A., and Kenneth W. Wachter. 2001. "Census Adjustment: Statistical Promise or Statistical Illusion?" Technical Report #596. Berkeley: Department of Statistics, University of California.

Navarro, A., and M. Asiala. 2001."Accuracy and Coverage Evaluation Survey: Comparing Accuracy." Report B-13. Washington, DC: U.S. Department of Commerce, Bureau of the Census.

Stark, P.B. 2001. Review of *Who Counts?* by Margo J. Anderson and Stephen E. Fienberg. *Journal of Economic Literature* 39:593–95.

U.S. Department of Commerce, Bureau of the Census. 2001. "Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy." Washington, DC: Department of Commerce.

Wachter, Kenneth W., and David A. Freedman. 2000. "The Fifth Cell." *Evaluation Review*:191–211.

Ylvisaker, Don. 2001. Review of *Who Counts?* by Margo J. Anderson and Stephen E. Fienberg. *Journal of the American Statistical Association* 96:340–41.

https://doi.org/10.1017/S1049096501000749 Published online by Cambridge University Press