Comment

Eugene P. Ericksen, Stephen E. Fienberg and Joseph B. Kadane

Our discussion of these three papers focuses primarily on those by Breiman and by Freedman and Wachter. Our observations are consistent with many of those made by Belin and Rolph in their paper, but this should surprise few people who are aware of both our previously stated positions on undercount adjustment and the role that two of us (Ericksen and Fienberg) had as expert witnesses in the recent New York City census adjustment litigation, described in part by Belin and Rolph.

1. BREIMAN ON BAD DATA

Until recently, many Census Bureau and other experts on census coverage equated the net undercount rate with the omission rate. For example, in a report describing the coverage of the 1970 census, the Bureau of the Census (1975) analyzed the consistency between the demographic estimate of undercount and the omission rate given by a postenumeration survey. They made no mention of the possible existence of erroneous enumerations, even though the bureau measured such errors as part of its first such survey in 1950.

In 1980, Bureau Director Vincent Barabba, in explaining his decision not to adjust the census gave as one of two reasons the fact that the net undercount was close to zero. Later analysis exposed the problem with his conclusion. First, the bureau reckoned that about 3 million undocumented aliens had been left out of the demographic estimate of the national population which was the basis of the conclusion that there was no undercount. Second, survey data collected in the 1980 Census Postenumeration Program indicated best estimates of 13 million omissions and 10 million counting errors, which are the sum of substitutions and erroneous enumerations. There appeared to be substantial variations in net undercount rates among places. The fact that the national estimates of omissions and counting errors were close was now seen as accidental.

In 1990, the situation was similar to that in 1980, but worse. The redefined question then asked by the Census Bureau was whether or not the observed geographic distributions of omissions and erroneous enumerations were real or were caused by errors in the data. In 1990, the estimated numbers of omissions (20 million) and counting errors (16 million) were much larger than they had been in 1980 (Bryant, 1993; Ericksen and DeFonso, 1993), and the net undercount (4 million) was slightly, but coincidentally, larger than it had been in 1980.

Given this background, the focus of Leo Breiman's paper seems misdirected. He concludes "The largest part of the original undercount estimate is due to bad data and processing error—80% on the national level." In Breiman's terms, he believes that the correct estimate of net undercount may be as low as 1 million. For this to be true, either the estimated number of omissions would have to be lowered from 20 to 17 million, the estimated number of counting errors would have to be increased from 16 to 19 million or there would have to be some combination of the two. Either way, there would be 30–40 million census errors to be accounted for, and if a decision was made not to adjust the census, one would simply have to hope that the distributions of these errors were so similar that between-area variations in net undercount rates would be minor. In our view, Breiman focused his time and energy on the wrong problem. Rather than trying to show how PES data problems inflated the national estimate of net undercount, he would have better spent his time showing how these errors might have skewed the estimated differentials between places.

Breiman's paper is based largely on the 1990 Post Enumeration Survey evaluation data, which came from three sources: (1) records of quality control procedures; (2) a repetition of matching procedures carried out for a sample of PES cases by more expert matchers at the Census Bureau; and (3) the Evaluation Followup Survey, in which a subsample of PES respondents were reinterviewed. This interviewing occurred in January 1991, fully nine months after census day and five to six months after the PES interviewing period. Using these evaluation data, Census Bureau statisticians had already assessed the quality of the PES data used for the undercount estimates. This evaluation was summarized by Mulry and Spencer (1993). In their best judgment, the national net undercount was *slightly* too high, but the differential undercount among places was substantially as the original PES had indicated.

Eugene P. Ericksen is Professor of Sociology and Statistics, Department of Sociology, Temple University, Philadelphia, Pennsylvania 19122. Stephen E. Fienberg and Joseph B. Kadane are Professors of Statistics and Social Sciences, Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213.