Rejoinder

Thomas R. Belin and John E. Rolph

NEW DEVELOPMENTS

Since the three articles and the discussions were accepted by *Statistical Science*, a federal appeals court set aside the judgment of a district court judge that had allowed the Commerce Department's decision not to adjust the 1990 decennial census to stand. Although the lower court's ruling was vacated, the ultimate consequences of the appellate decision are far from clear. Actions by the plaintiffs, by the government, by other appellants and ultimately by the courts will determine how the 1990 census adjustment saga is played out. See Fienberg (1994b) for a more detailed discussion of the appeals court ruling.

We hesitate to read the appeals court ruling as an endorsement of our scientific point of view. Nevertheless, the rulings of both the district court and the appeals court reflect a willingness on the part of nonstatisticians to view an adjusted census as a feasible and reasonable approach for improving on the accuracy of an attempted headcount.

Statisticians should understand and appreciate this willingness. We thus hope that this appellate court decision will give a new impetus to the statistics community to help facilitate consensus on how to use estimation methods in census-taking. The Census Bureau's investigation of a "one-number" census is a constructive step in this direction.

RESPONSES TO DISCUSSANTS

We focus here on the discussions by Diamond and Skinner, by Steel, by Lyberg and Lundstrom, and by Ericksen, Fienberg and Kadane; we have not seen either the Freedman and Wachter (FW) or the Breiman rejoinders, although we comment briefly on a point raised in some exchanges with our Berkeley colleagues.

The discussions by Diamond and Skinner, by Steel, and by Lyberg and Lundstrom all provide useful and enlightening perspectives on census-taking practices around the world. The balance in their remarks sets a good example for us to follow here in the United States.

The final paragraph of the discussion by Diamond and Skinner amounts to an excellent summary of our essential points: the debate over census adjustment should emphasize scientific matters, but consensus will require more than just scientific progress. We appreciate their supportive remarks. On a more subtle matter, their point is well taken that the term "heterogeneity bias" could be construed to mean either error in the synthetic assumption explored by FW or error in the assumption of equal capture probabilities discussed, for example, in Alho, Mulry, Wurdeman and Kim (1993). We hope our use of the term was clear from the context. We are also glad that Diamond and Skinner agree with us that the heterogeneity reported by FW is not surprising and that heterogeneity should be reflected in local-area estimates of variability.

Steel's discussion of how statistical estimation is used in the Australian census provides a valuable frame of reference for the debate about the future of the U.S. census. Steel does not attempt to adjudicate our disputes with Breiman, yet we interpret his remarks in his penultimate paragraph as supportive of our point of view: the census-taking process has to stop somewhere, and decisions have to be made. We read Steel's final paragraph as reflecting a semantic difference with our use of the term "consensus." We do not imagine that all statisticians would realistically line up behind one particular census methodology, but we can imagine there being a critical mass of support for a particular approach so that the controversy subsides. Indeed, during the 1980's as a member of the National Academy Census Panel, one of us (Rolph) saw such a critical mass forming on that panel and among the senior staff of the Census Bureau behind a planned adjustment methodology for the 1990 census. Intervening events led to the controversy and adversarial process referred to in these articles, but in our view the process need not be so contentious. For example, the use of postal delivery in the 1960 census was a major methodological change, but one that did not engender much controversy. There may have been some people who opposed this innovation at the time, but the level of support for a mail-out, mail-back census would qualify as consensus from our standpoint.

Lyberg and Lundstrom add a variety of insights that reflect the realities of government statistics practice. In a few places, their statements are stronger than we would make. For example, we do not have a problem with criticizing "bad data"; what we object to is the notion that we should assign a loss of infinity to model-based estimators and then call such an approach good science. We are also

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *Statistical Science*. www.jstor.org

more optimistic than Lyberg and Lundstrom that further investment in methodology will pay dividends. One example of methodology that we cited repeatedly in our paper is Zaslavsky (1993a); this work not only advances the state of the art for census undercount estimation, but it serves as a useful case study that could be adapted to other statistical arenas as well. Overall, however, we appreciate their endorsement of our general perspective on the adjustment controversy.

Ericksen, Fienberg and Kadane make few comments directly about our paper. We would simply point out that some of their recent references (Kadane, Meyer and Tukey, 1992; Darroch, Fienberg, Glonek and Junker, 1993) also serve to illustrate that progress is still being made on undercount-related issues, yielding both new theory and new methods.

We understand that in their rejoinder, FW cite a personal communication from us. We offer the following comment in the spirit of "setting the record straight."

In the initial version of his paper, Breiman made a stronger claim about the increasing proportion of unresolved cases in the Evaluation Followup Survey (EFU) when one reads across his Table 12, which we saw as the kind of nitpicking criticism that deserves to be pushed to the margin. Originally, after we pointed to Breiman's curious claim that one might do just as well in imputing for unresolved cases by flipping a coin with probability 15% of heads, we had written, "The higher proportion of remaining unresolved cases in the higher imputed probability categories is explained in large part by the fact that names were not recorded for many PES individuals." However, it turned out that our explanation was inaccurate; although cases without names constituted approximately 70% of the P-sample cases receiving probabilities of having been enumerated of 75–100%, these cases without names were largely excluded from the EFU and so were not reflected in Breiman's Table 12.

We acknowledge that there were a substantial number of unresolved cases in the EFU and that there is remaining uncertainty about the accuracy of the imputation methods. Our essential point is that there is not much to criticize based on available data, which agree with predicted values extremely well (Belin et al., 1993). To attribute our earlier statement to us as if it is our current view is a misrepresentation.

Overall, although we anticipate that our Berkeley colleagues will continue to support one another, we are pleased at the signs of consensus in this exchange.

Rejoinder

Leo Breiman

I thank the discussants. The descriptions of the methods used in Australia, Great Britain and Sweden were interesting and form a compact introduction to the diversity of methods for estimating population counts. They also underline the difficulty of the census undertaking in the United States. The discussion by Ericksen, Fienberg and Kadane and the Belin–Rolph article contain most of the direct comments about my paper.

BACKGROUND

The effort to adjust the census counts was a complex process. After the initial error evaluation, additional errors were discovered, some of which are discussed in my article. Because the original error analysis has not been updated to take these additional errors into account, the widespread impression remains that the adjustment process was proven to produce more accurate counts than the census.

The validity of any such proof is currently in serious doubt. For one thing, errors of various types are now acknowledged to account for the major part of the original national undercount estimate of 2.1%. The initial loss function analysis used earlier estimates of the bias that, on the national level, were too small by at least a factor of 2. The analysis was also flawed by a significant underestimation of sampling variances (Fay and Thompson, 1993; Freedman, Wachter, Cutler and Klein, 1994). There are also questions about the additional local bias due to heterogeneity (Freedman and Wachter, 1994), the errors resulting from smoothing the adjustment factors (Freedman et al., 1993) and many of the assumptions going into the loss function analysis (Freedman, Wachter, Cutler and Klein, 1994).

This careful scrutiny was possible, in part, due to the availability of three sets of numbers: the census counts, the adjustments and the extensive evaluation data. We view the controversy over