DISCUSSION

Michael A. Stoto, National Academy of Sciences
2101 Constitution Ave. N.W., Washington, D.C. 20418

KEY WORDS: census undercount, census adjustment, post enumeration surveys, dual system estimation

I am pleased to have the opportunity to discuss these five interesting and exciting papers. Together they show how far we have come in undercount/adjustment research, despite considerable adversity. I will try to address some comments to each paper individually, as well as some comparative and overarching issues.

Hogan paper

Hogan begins his paper by asking: "Nine years of coverage evaluation research: what have we learned?" Based on the rest of the paper, the answer is clearly: "A great deal!"

I found the historical review to be very interesting. It contains much information that I wasn't aware of, despite following this issue carefully for almost a decade. It also sets the stage for understanding subsequent decisions and actions. From a historical viewpoint, however, it would have been good to have more individuals' names in the paper. The current paper gives the impression that important decisions were somehow just made, when in reality nearly everything must have been the product of much controversy and strong personalities. Many of the people who made those decisions are still around and still involved in various ways. Knowing their personal history on this issue would help us to understand their current positions.

Overall, Hogan's paper and the others in this session show what can be accomplished by persistence. Hogan makes the case that people have been working on the census undercount issue for half a century. There were many apparent failures along the way, but the cumulative, collective knowledge about census undercount and adjustment issues is far advanced.

More specifically, the paper shows the value of labeling and categorizing undercount estimation techniques, sources of error, and so on. These are the vehicles for building collective knowledge and for getting new people into the field without having to learn everything from scratch. Labels and categories are especially important for a subject, like the undercount, where both the operational and theoretical issues are complicated and complex.

To appreciate the papers in this session, it is important to know the status of census undercount and adjustment issues. In 1987, The Department of Commerce announced that the government was not going to even consider adjusting the Census, despite the research progress described by Hogan and plans for a careful and timely estimate and despite the advice of many of the Bureau's technical advisors. Last year, New York City and others brought suit against the Department to reverse this decision. Then just last month, both parties agreed in an out-of-court settlement to (1) carry out a large-scale post enumeration survey in a timely fashion, (2) publish a set of guidelines to guide the adjustment decision before the census is taken, (3) attempt to develop adjusted figures by July 15, 1991, and either publish the adjusted figures or an explanation of why an adjustment was not feasible, and (4) set up a panel of experts to advise the Bureau on these matters.

This settlement breathes new life into the undercount/adjustment research area. The decision process is very much like the one that was being discussed five years ago. The main difference is that some momentum was lost for 18 months or so following Commerce's 1987 announcement, especially in terms of people and research. On the other hand, the Bureau now has an additional 6.5 months (from December 31, 1990 to July 15, 1991) to consider and to carry out an adjustment. As Hogan says, what happens in the next two years (specifically up to July 15, 1991) will be a real test of the research in the last decade and before.

A major part of the settlement calls for the development and adoption of a set of guidelines to articulate what the Bureau believes are "the relevant technical and nontechnical statistical and policy grounds for decision on whether to adjust the 1990 Decennial Census population counts." These guidelines are to be published for comment by December 10, 1989, and in final form by March 10, 1990. They will draw heavily on the research that Hogan described, especially on the categories of errors to be looked for and on the relationship between observable variables, such as response rates, and overall accuracy. The comment period this winter will be an opportunity for all statisticians to participate in this great national experiment, and I urge you all to participate

Mulry/Dajani paper

The Mulry/Dajani paper is a good example of the many special studies in the undercount research program in the last nine years. It is a very practical study with an important and well defined purpose.

The paper makes a good case that (1) 5 to 10 percent of people counted in the census could not be traced for five years; (2) substantially higher proportions (up to 35 percent) of people who were missed, born or moved into the country in the intervening years can't be traced; (3) there were also differences by race; (4) different contact patterns and levels of effort make a difference, but not much difference; and (5) the operation was difficult to control, and would be difficult to carry out on a large scale at census time.

Mulry and Dajani then draw the implications that the situation would be much worse for a ten-year intercensal interval. Common sense tells us that this is true, but it is not tested in the paper. It would have been good, for instance, to see how the trace rate deteriorates over the five years for which data are available. Mulry and Dajani also seems to suggest (but do not say outright) that the attainable trace rates are too low for the method to yield useful undercount estimates.

I see two problems with this paper. First, the records do not seem to have been kept in a way that allows the authors to estimate costs of the treatments or techniques. This should have been designed in from the beginning. It is difficult to know what to make of the cost estimates that are available. The estimated cost for tracing one person ranges from \$30 to \$100 for people enumerated in the last census depending on the level of effort. Given these numbers, is it worth it to increase the level of effort? How would you know?

The other problem is the evaluation framework. To be fair, this paper seems to be focused on the operational side of the forward trace, not on the ability of the resulting data to improve undercount estimation. It would be nice, however, to try to take the analysis one step further and work out something about the quality of the undercount estimates that could be expected. To do so, one would have to think about how the data would be used, and the only way I know to do this is through a dual-system estimation approach. This suggests that you don't need full coverage for the estimation to work, just independence between the forward trace and the census. Could racial differences in trace rates be taken into account by disaggregation? Does the difference between the C and M sample trace rates translate into an important correlation bias?

Wolfgang and Zaslavsky papers

This is a very interesting pair of papers. Both address the general issue of what to do when you have more than two lists (census, PES, and others), but the authors take very different approaches.

Wolfgang's paper is very practical, dealing with an actual exercise carried out in conjunction with the 1988 PES test. He finds that additional men can be found by culling through a variety of administrative lists. This technique increases the estimate of the undercount in places where we suspect that correlation bias would lead the PES to underestimate it, and makes the sex ratio closer to what we would expect it to be. Note, however, that the people doing this project had to cull through nearly a million records from six separate sources to find an additional 349 cases.

I have three questions about this paper. First, why were new cases simply added to the PES sample rather than entered into some kind of dual-system estimation? I think that Wolfgang's assumption is that the additional lists will make the PES sample almost complete, therefore denying the possibility for correlation bias. It seems as if the methods discussed and developed by Zaslavsky would have been appropriate and informative here.

Second, I would like to hear some more on the operational feasibility of doing this operation on more than a few PES sample blocks.

Third, is it appropriate to carry out this type of operation is some strata but not others? If so, which ones?

In contrast, Zaslavsky takes a theoretical approach, developing models for incorporating data from multiple population lists. His paper provides a framework and language for thinking about multiple-system estimation, using both Bayesian and empirical Bayesian estimation, and especially the problem of correlation bias. He also shows quite clearly how to deal with data that is available for nested and non-nested subsets of the population.

I have two questions about this paper as well. First, what happens if different kinds of lists are available in each place, or if the quality of data varies from place to place? For example, suppose driver's license data is available in one area, IRS data in another, and both in a third? Presumably it is like Zaslavsky's model 2c or 2d, but at what number of different kind of lists does it break down? Or, suppose different amounts or kinds of matching data are available on each list. Must each list be treated as a separate source

Second, how can this model incorporate the kinds of missing and incomplete data problems that actually occur in carrying out the PES field work? As Schafer's paper (and the 1980 New York suit) make clear, treatment of these issues matters a lot.

Schafer paper

This very interesting paper that actually goes far beyond what the title, "Why P-sample Set 1Y Should Not Be Trusted", implies. It provides a general framework for thinking about imputation/reweighting and treatment of missing data in the PES. This issue was a critical part of the bureau's argument in the 1980 New York case. Schafer's results should help in developing the 1990 analysis plan.

Schafer's is an excellent example of how careful description of a process in mathematical terms and generalization can clarify a complicated issue. Although the choice seems obvious after reading this paper, many smart people were very confused about this in the early 1980s.

Conclusion

In conclusion, I would like to thank the speakers for an excellent and enlightening set of papers. Together, they indicate substantial progress on an important and difficult statistical policy problem. The Bureau should also be congratulated for their support of this research by their own staff and others they have funded. It is a credit to their commitment to quality and objectivity in national statistics.