DISCUSSION

Benjamin King, Florida Atlantic University
220 S.E. 2nd Avenue, Ft. Lauderdale, FL 33301

It is a privilege to be able to discuss these interesting and informative papers. Since this is the last in a whole series of sessions at this meeting devoted to the 1990 census undercount and the post-enumeration survey, and with the decision by the Department of Commerce against adjustment in mind, it occurred to me this morning that this session ought to be entitled either Diehard IV or Diehard VI, depending on whether or not you count the invited paper sessions.[1]

The first four papers and the sixth are reports of results from PES evaluation studies. I shall discuss them all before saying a few words about Zaslavsky's work, and in so doing change their order somewhat from that in which they were presented. In the short time available for each I shall try to comment on what I see as the principal information that they give concerning the reliability of the PES and its effect on undercount estimates.

1. First the paper by Anolik and Hogan on nonresponse conversion: The low percentage of PES noninterviews was impressive by any survey research standard, and I was especially pleased to learn that "last resort" cases were treated as noninterviews. The Nonresponse Conversion Operation (NRCO) was aimed at gaining further completed interviews and match status. This is not perfectly clear from the paper, but I assume that the results of that operation were added to the original PES and used in the final dual system estimation.

Considering those converted in NRCO to be relatively "hard core" cases, we can see from Tables 2 and 3 how these persons compare with those who cooperated earlier. They are not very different, although there are some points of difference worth noting. For example, household size is smaller for the NRCO respondents-- as expected. The text states also that there are relatively more blacks in the NRCO group, and this is explained as a result of the district offices in NRCO being in larger urban areas. In the version of the paper that I read, however, Table 3 showed that there are relatively fewer blacks among the NRCO respondents. The table was mislabelled and subsequently corrected by the authors, but not knowing which statement was true, I explained the lower representation of blacks as indicative of greater tenacity in noncooperation by the minority group. I mention this point of confusion because it just goes to show that almost any result that one finds can have a ready explanation that is compelling and logical.

The most important finding of this research is that the match rate for the harder-to-convert respondents is lower than for the original response group, implying that those who were not counted in the census tend not to get enumerated in the PES-- evidence of correlation bias. Assuming that the match rate is even lower for those who are still holding out, this indicates that the undercount rate is underestimated. This is also shown in general, by the results in Table 5, comparing undercount estimates with and without NRCO interviews. I find this to be reassuring in that it confirms the claim made all along that PES results would be conservative.

2. The leadoff paper by Philip Gbur is entitled "Missing Data", and it also gives some information on rates of nonresponse, based on the relatively small scale evaluation sample. I must confess, however, to some confusion over how to use that information on noninterview rates in assessing the adequacy of the PES. The most interesting results, to me, are (1) those dealing with the comparison of imputation rates and estimated undercounts across post-evaluation strata, and (2) the comparison of imputed match probabilities with actual match statuses determined for converted noninterviews. Gbur explains that there are two alternative reasons for the moderately high positive correlations between the estimated percent undercount and the percentage of imputed characteristics. Either areas with truly high undercounts tend to have more missing data, or else the imputation method is causing the high undercount estimates-- possibly because of errors of underestimation in the numbers of matches in areas with a lot of missing data. I think that the issue is cleared up pretty well by Table 9 in which we see that when converted and resolved cases in the evaluation sample are classified by their imputed probabilities of a match, the relative frequencies of match and nonmatch are in good agreement with the prediction. This increases my confidence in the match probability imputation model.

3. The paper by Alberti and Anolik, "Matching Movers in the 1990 Post-enumeration Survey", tells us all that we ever wanted to know about persons who moved between Census Day and their interview in the PES. It is a very thorough and commendable report. I believe that, because of its clarity of exposition, I understand everything that it is saying, and I have no particular comments except to lament the fact that we could not keep people locked in their houses during a period starting before the census and ending after the PES. Movers appear to me to be the most serious potential source of error in coverage evaluation. I guess that the best way to attack the problem is to figure out how to start the PES much sooner after the census. I have not kept up with the current discussions on innovation in the Year 2000 Census, but I certainly hope that the planners will consider the coverage evaluation activities to be as much an integral part of the operations as the main count itself. I have a feeling that even though there will not be an adjustment this time, things will never be quite the same again and that we are clearly converging on well-founded adjustment someday not too deeply into the twenty-first century. At any rate, I call for a PES that contacts each sampled housing unit as soon as the census taker has walked out the door, and with a pontifical wave of the hand I leave it for the great minds at the Bureau to come up with a suitable high-technology solution to the chaos that would ensue.

4. If a mover is not matched (either by actual determination or through imputation), that person contributes to the estimated undercount in the poststratum containing the census-day location. Schindler and Griffin, in their paper, "Treatment of Differential Weights...", consider the circumstances where a mover receives a certain estimation weight in the PES and is moved back into a poststratum where the weights of the other individuals are of different magnitude. As is well known, if the range between the highest and lowest weight in a domain is great, and the shape of the distribution of weights is unfavorable, then variances of domain estimates can be very large. The authors have examined through case analysis and simulation the desirability of "weight capping", i.e., restricting the maximum weight to a certain limit, thus accepting bias in exchange for variance reduction. They conclude that capping

does not buy enough to make it worthwhile. Since decisions about weighting had to be made in a timely manner this analysis was carried out before the PES data were available for detailed evaluation. Modification of the parameters of the analysis based on PES results have not changed the conclusions.

My only suggestion-- for the sake of making some minor critical comment-- is to wonder how much the variance estimates would change if the weights were not considered fixed, but rather, treated as random variables depending on the rates of nonresponse. My guess is that they would not be affected very much. As for the importance of this issue for the overall reliability of the undercount figures, matching errors among the movers are a much more serious matter than the effects of differential weighting, which is not to say that this investigation was not a necessary step in the overall evaluation.

5. I enjoyed reading the paper by Mack, "An Analysis of Reasonable Imputation Alternatives...", (the second in order of presentation). The experiments described therein are very elegant. The analyses show that changing the method of imputation for unresolved matches, and especially considering proxy reports to be unresolved, does not introduce variability into undercount estimates nearly as great as sampling error itself. The analysis of the bootstrap data similarly shows that the effect of variability in estimation of logistic regression parameters is negligible. My one question concerns the sampling standard deviations used in Figures 2 and 3. I assume that they are the single estimates that result from the production run, or are they an average of sampling error estimates under each of the treatment combinations?

6. Now a few words about the Zaslavsky paper, "Combining Census and Dual-system Estimates...". It seemed natural to save this for last because of its global orientation. It describes an ambitious and courageous attempt at modelling the whole process of dual system estimation and evaluation, including the choice of best estimator through minimization of expected loss. About the only element in the global specification that is left out is the Department of Commerce, itself -- we can leave the modelling of that for a doctoral dissertation in political science.

The paper describes two main activities: The first is a set ofsimulation runs in which prior parameters are varied systematically and random trials generated according to the selected parameters. The performances of alternative methods of estimation are compared. It is not surprising that Hierarchical Bayes 3 performs well, since it takes bias into account more explicitly than alternatives. As with most simulations, I am left wondering how realistic the chosen ranges of parameters are, and what the chances of encountering the worst case scenarios are in practice. This might be called the "fundamental problem" in simulation studies.

The second set of analyses is perhaps the more useful because it incorporates recent estimates of variance components and biases from the PES evaluation program. It is intended to check on the sensitivity of the relative performances of the Bayes 3: HB estimator to changes in the parameters that determine error in the dual-system, distinguishing between those that potentially can be estimated from the data (albeit not easily) and the parameters assumed in the modelling, but not verifiable under present evaluation methods. You have just heard the author report that the insensitivity of the estimates to the choice of assumptions is remarkable.

In closing, I just want to say that although this full-blown Bayesian examination of the total system is in many ways mind-boggling, it is, in my opinion, the only way to go. Yesterday morning we heard Bruce Spencer describle his study of total error as valuable because it exposes assumptions and helps to focus debate. In the same way, Zaslavsky's work forces us to consider things that have to be considered, as painful and complex as it may be. Yesterday we also heard the work of Mulry and Spencer described as "story-telling", and not statistical science as it should be practiced. Well, if Mulry and Spencer is story-telling, then Zaslavsky's paper is a veritable Bulfinch's Mythology of statistics. To those critics I say that blind acceptance of the status quo and nihilism hardly serve well as paragons of good statistical science.

For Zaslavsky's work, and for the work of the other authors in this whole effort of the PES and its evaluation that we have been hearing about these last few days, I call for a hearty, "Hip, hip, hooray!"

1. For the information of those historians who may be reading this in the twenty-first century or beyond, Diehard refers to a recent popular film and its sequel-- a feeble attempt at levity after an arduous four days and nights of statistical meetings.