DISCUSSION

Robert E. Fay, U.S. Bureau of the Census

The title of this session, "Problems of Nonsampling Error in Surveys," conveys aptly the common theme of these papers. Two papers concern nonresponse, one presents statistical methods for the analysis of interviewer variability, and one examines internal evidence and compares survey results with independent estimates to evaluate the effect of nonsampling error on a survey. All are interesting and instructive.

The paper of Phillip S. Kott examines the problem of imputing missing data in a time series. The paper explores alternative imputation procedures and considerations in choosing among them. Specific statistical tests are proposed to assist in the selection.

The principal focus of the paper is the impact of the imputation procedure upon the variance of the final estimate. This is an important issue and one that may be too frequently overlooked in practice. In making a choice of imputation procedure, however, other factors are also important. Specifically, the imputation procedures examined in the paper make different assumptions about the missing data, and the appropriateness of the respective sets of assumptions is often a more important issue than variance. All of the alternatives provide a large-sample consistency if the data are missing completely at random in the sense of Rubin (1976), that is, missing with the same probability regardless of the observed or unobserved values of the survey variables. This assumption is generally quite severe, however, and is refuted by the data whenever nonresponse rates vary systematically with observed characteristics. The different imputation procedures examined in the paper allow the assumption that the data are missing completely at random to be relaxed, but in different ways.

The paper appears to contrast the modeling of response behavior with the modeling of "parametric behavior." The author's statements on this subject seem to favor the latter approach over the former, and his paper does not elaborate any response models beyond the assumption that the data are missing completely at random. Nonresponse is clearly an issue of individual behavior, however, and explicit models for the propensity to respond are the most effective means to elicit the assumptions underlying any approach to missing data, in my opinion. The paper would have benefitted from greater focus on this aspect of missing data.

One of the first problems that the paper addresses is whether mean imputation or the ratio-of-identicals method yields lower variances. This problem exactly parallels the question "Under what circumstances should one use a ratio estimator, rather than an expansion (Horwitz-Thompson) estimator?" Answers to this second question can be found in the traditional literature on survey sampling. More precisely, each of the two estimators may be characterized as a two-step procedure, in which an estimate of total for the n sample observations is constructed in the first step and inflated to an estimate of total for the population of N elements through simple expansion by the factor, (N/n), in the second step. Mean imputation is equivalent to use of a simple expansion estimator at the first step as well, while the ratio-of-identicals method uses the classical ratio estimator for this first step. Since both procedures employ the simple expansion estimator at the second step, the question of which imputation procedure yields the lower variance is thus equivalent to ratio vs. expansion estimation. Theorem 1 of the paper may be derived by direct application of eq. (6.5.11) of Kish (1965, p. 204) or Theorem 6.3 of Cochran (1963, p. 165). Although the result stated in the paper is correct, errors in the derivation appeared in the version of the paper available for my review.

The last part of the paper develops a time series approach to the imputation problem, but, unless the time series were quite long, e.g. 50-100 observations, I would prefer to view the problem in the context of a series of linear regression equations. The dependent variable would be the current value, and previous values could be used as the predictors. Some transformation of the data, particularly the log transformation, may may give a more suitable functional form. Depending upon how complex an approach seems warranted, the E-M algorithm may be required to estimate the equations in the presence of missing data.

Lynn Stokes and Joe Hill nicely illustrate the application of two statistical methods to the analysis of data from interviewer variance studies: empirical Bayes estimation and generalized linear models. Their work should be especially stimulating to researchers concerned with the measurement of interviewer quality and consistency.

I have a few misgivings on actual application of these methods to improve the estimation of the population mean, however, in place of estimators based upon traditional sampling theory for finite populations. First, users of these methods in other settings should be quite careful to avoid possible time-of-day effects that could arise if some interviewers worked at different times than others. Secondly, the improved estimator of the mean is based upon "down-weighting" the results from interviewers with larger production. In some instances, such interviewers may be more experienced and possibly subject to lower interviewer variance than those with the least output, who may be newly hired. In such circumstances, the reweighting of the survey data to favor those with lower outputs may be disadvantageous in terms of overall quality. I hope that the authors are able to pursue further research on the robustness of their model.

The paper by Elizabeth Stasny describes a number of models for nonignorable response, especially of panel data. This paper complements the first paper, by considering the issue of modeling the mechanisms of nonresponse, although the focus here is categorical rather than continuous variables. The models discussed should be of great interest to those who analyze panel data subject to nonresponse.

Development of nonignorable models for categorical data has been an active area of research recently. In addition to her own work and the references cited by her, other manuscripts in this area have been prepared; some have received only a limited, informal circulation. Specifically, a

paper by Baker and Laird (1985) discusses models for a single variable subject to nonresponse, and a paper of Little (1985) examines such models and reviews other work in this area. Another manuscript (Fay 1985) characterizes a class of models for nonignorable nonresponse for categorical data by recognizing the close connection to work of Leo Goodman on causal models for categorical data. This class of models includes those considered in the paper by Stasny. A benefit of recognizing the connections among the specific models of this class is that a common approach based upon the E-M algorithm facilitates estimation. The paper by Stasny makes an important contribution by discussing and illustrating the application of these models in the setting of panel data.

Although nonresponse is an important issue in interpreting the data on gross flows from the Current Population Survey, the available evidence suggests that an even more important practical issue arises from the large overestimation of the month-to-month change arising from response variability in the survey data. This issue does not detract from the methodological interest of the paper, but this limitation of the CPS data nonetheless deserves mention.

Charles W. Warren presents a detailed evaluation of the data from the 1982 Puerto Rico Fertility and Family Planning Assessment (PRFFPA). His paper makes a useful contribution generally, and certainly should be of great importance to those analyzing this particular data set.

The principal method of evaluation in his paper is one that is an apparent favorite of demographers - the analysis of aggregates. Perhaps the most salient feature of this method is the very high benefit/cost ratio, where much can be learned about a specific data set by careful examination of the results. Aggregate comparisons such as those appearing in his paper also serve the function of clarifying the degree of consistency between different important sets of data.

Some limitations of this method also deserve mention, however. Generally, the actual reliability of the data is not entirely clear from comparisons of aggregates. For example, some care must be taken in interpreting an agreement between census and PRFFPA as a guarantee that both were correct.

Table 2 of the paper, which shows characteristics of nonrespondents, merits more thorough study. In spite of the size of the sample, the response effects shown there are quite considerable and help to explain some features of the data. Because the respondents are disproportionately married women, fairly significant effects on estimated fertility from PRFFPA may be hypothesized.

The paper could have benefitted from a more systematic treatment of the issue of sampling error. The estimated value of Myer's Blended Index suffers from substantial bias arising from sampling variability, which could be largely removed through use of a jackknife or other suitable replication techniques. Several inferences in the paper could have been strengthened or better substantiated if sampling errors had been available.

The effect of a possible undercount in the census has considerable implications for any comparisons of the survey to census aggregates. In particular, the possible effect of census undercount must be taken into account in comparisons between the census and vital statistics.

References

Baker, S. and Laird, N. (1985), "Categorical Response Subject to Nonresponse," working paper, Department of Biostatistics, Harvard School of Public Health.

Cochran, W. G. (1963), Sampling Techniques, New York, NY: John Wiley & Sons.

Fay, R. E. (1985), "Causal Models For Patterns of Nonresponse," submitted to the Journal of the American Statistical Association.

Kish, L. (1965), Survey Sampling, New York, NY: John Wiley & Sons.

Little, R.J.A. (1985), "Nonresponse Adjustments in Longitudinal Surveys: Models for Categorical Data," presented at the 1985 Meetings of the International Statistical Institute, Amsterdam.

Rubin, D.B. (1976), "Inference and Missing Data," Biometrika, 63, 581-592.