DISCUSSION

D. G. Horvitz, Research Triangle Institute

It is certainly heartening to attend a session where the focus is on non-sampling errors, particularly their sources, how to estimate their identified components and how to reduce or otherwise control their contribution to the total error of sample survey estimates. While progress has been made in the non-sampling error struggle, much much more remains to be done.

Norman Beller certainly makes a telling point when he raises the paradox of the increase in non-sampling error realized by the use of complex survey designs originally intended to reduce the sampling error. Clearly, more and more attention needs to be given to the use of total error models and to appropriate cost functions for complex survey designs in order to determine optimum trade-offs. Beller points out that reducing the list sample may result in a reduction in the non-sampling error. This suggests to me that the relationship of non-sampling errors to size of list farm should be investigated to test the assumption that domain classification is more accurate for larger farms. If true, confining the list sample to the larger farms could reduce domain errors.

Some alternative to total farm land may be more appropriate for the weighted segment estimator. For example, acres of cropland harvested may be estimated by farmers better than total farmland. A study by Fleischer et al. (1958) indicates that farmer reports of sizes of cultivated fields are unbiased. Cropland harvested may also be somewhat more stable over time than farmland (Hendricks et al., 1965). Some other weighting variable such as grazing land may be more appropriate, of course, for cattle and hogs.

To me, the most revealing aspect of the Krotki paper is the instability of the estimates of the correlated response variance (CRV). The coefficients of variation are often greater than 100 percent. Thus, there is no basis for deciding that the new CRV estimator is any better than the old, or even that the combined estimator is better than either the old or the new. The problem is clearly one of either developing an improved estimation technique or of increasing the number and amount of interpenetration; that is, the effective sample size. It may be better, for example, to divide each EA into four equal sized groups and assign four interviewers to one each of these groups at random in four different EA's than to have two EA pairs and use only two interviewers per pair.

Paul Biemer's paper is highly correlated with Karol Krotki's in that both seek more efficient estimators of the interviewer contribution to the total variance and both use non-interpenetrated interviewer assignments as well as interpenetrated assignments to achieve an increase in efficiency. It would be useful to know whether Biemer's design and estimating procedures are, in fact, more efficient than Krotki's.

Hartley and Monroe seek to optimize interviewer assignments in order to minimize the correlated response variance. In the process they point out that interviewer assignment schemes that provide more within stratum interviewer contrasts should also provide more reliable estimates of the interviewer variance component. Unfortunately they are not able to quantify the gain in efficiency achieved through additional interpenetration.

Some attention must be given to the trade-off in terms of costs and efficiency between using the Hartley and Monroe optimum interviewer assignment design and increasing the number of interviewers, but using current conventional assignment schemes. It should be noted that doubling the number of interviewers (2k) would achieve a slightly greater reduction in the interviewer variance contribution to the total variance than would the Hartley and Monroe optimum interviewer assignment design with k interviewers.

It is not clear to me why Biemer is concerned with designs which include coder as well as interviewer assignments. Since quality control procedures can be designed to control coding errors at a specified level, whereas quality control in the same sense cannot be imposed on the interviewing process, I would think the coder error variance would be orders of magnitude smaller than the interviewer error variance.

Both the Hartley and Monroe paper and the paper by Paul Biemer use an extension of an additive error model proposed by Hartley and Rao (1978) in order to derive estimates from survey data of the total variance of linear estimators of population parameters. This model is also used by Hartley and Biemer (1978). These authors are aware of the Bureau of Census survey error model originally developed by Hansen et al. (1961) for dichotomous variables and simple random sampling. They are not aware of the completely general extension of the Bureau of Census model by Koch (1973) to the multivariate case and for continuous as well as qualitative variables. Koch considers error variance components arising from the interaction of sampling and response errors as well as the usual variance components such as the pure sampling variance, the simple response variance and the correlated response variance. The immediate value of Koch's extension is twofold: first, it includes complex bivariate estimators such as ratio estimation and regression and correlation coefficients; second, it is not confined to simple random sampling, but may be applied to multistage clustered unequal probability sampling designs.

I suspect that the error models used by Hartley and Monroe and by Biemer in their papers are essentially special cases of the Koch model. If not, some discussion of the distinguishing features and their advantages would be helpful.

The Koch model served as the basis for a group of reports prepared in 1972 by Gad Nathan, Gary Koch and Babu Shah at the Research Triangle Institute under contract with the Bureau of the Census. These reports were concerned with estimating the sampling, simple and correlated response components of the overall variance of ratio estimates, and of sample regression and correlation coefficients; with survey designs for estimating response error model components; and with the stability of estimators for response error components. Copies of these reports should be available through Barbara Bailar at the Research Center for Measurement Methods, Bureau of the Census.

References

Fleischer, Jack et al. (1958). Measurement Errors Associated With Obtaining Acreage Estimates of Cotton Fields, Biometrics, 14, 401-407.

Hansen, M. H. et al. (1961). Measurement Errors in Censuses and Surveys, Bulletin of the International Statistical Institute, 38, 359-374.

Hartley, H. O. and Biemer, Paul P. (1978). The Estimation of Non-Sampling Variances in Current Surveys, Proceedings of the Section on Survey Research Methods of the American Statistical Association, 257-262.

Hartley, H. O. and Rao, J. N. K. (1978). The Estimation of Non-Sampling Variance and Components in Sample Surveys, Survey Sampling and Measurement, Academic Press, New York.

Hendricks, W. A. et al. (1965). A Comparison of Three Rules for Associating Farms and Farmland with Sample Area Segments in Agricultural Surveys, Estimation of Areas in Agricultural Statistics, Food and Agriculture Organization of the United Nations, Rome.

Koch, G. G. (1973). An Alternative Approach to Multivariate Response Error Models for Sample Survey Data with Application to Estimators Involving Subclass Means, JASA 68, 906-913.

Koch, G. G. (1973). "Some Survey Designs for Estimating Response Error Model Components," Technical Report No. 5, Contract No. 2-35018, prepared for U.S. Bureau of the Census by Research Triangle Institute, Research Triangle Park, N.C.

Nathan, G. (1972). "The Estimation of Correlated Components of Response Errors from an Enumerator Variance Study," Technical Report No. 1, Contract No. 2-35018, prepared by U.S. Bureau of the Census by Research Triangle Institute, Research Triangle Park, N.C.

Nathan, G. (1972). "The Estimation of Response and Sampling Error Components of a Ratio," Technical Report No. 3, Contract No. 2-35018, prepared for U.S. Bureau of the Census by Research Triangle Institute, Research Triangle Park, N.C.

Nathan, G. (1972). "The Estimation of Response and Sampling Error Components of the Sample Regression Coefficient," Technical Report No. 4, Contract No. 2-35018, prepared for U.S. Bureau of the Census by Research Triangle Institute, Research Triangle Park, N.C.

Shah, B. V. (1973). "A Stability Study of Estimators for Response Error Components," Technical Report No. 6, Contract No. 2-35018, prepared for U.S. Bureau of the Census by Research Triangle Institute, Research Triangle Park, N.C.