DISCUSSION ON SAMPLING FRAMES

Ron Fecso, National Agricultural Statistics Service, USDA
USDA/NASS, Room 305, 3251 Old Lee Hwy., Fairfax, Va. 22030

The papers in this session present many inter-related issues in the selection, creation and use of sampling frames. Sampling frames are addressed as not only the collection of sample units, but also the associated auxiliary information useful for design choices and the association rules necessary to link a respondent with the unit. While list frame issues (all five papers) and estimation (four) dominate the session, area frame, frame quality, completeness, burden, and association rules are also covered.

Much of the quality of a survey depends on the decisions made in frame development. Frame activities contribute considerably to the cost of many surveys. Yet, few opportunities exist for formal training in frame development. Since many of us learned much of what we know about frame development from proceedings papers, this session makes a valuable contribution.

Reading the papers inspired me to revisit one of my visual aids related to how we spend our time working on surveys. Frame decisions influence the choice of respondent, mode of interview, selection techniques, interviewer selection and training needs. Thus, the survey pyramid (Fecso, et al. 1996) needed a revision to reflect the importance of frames considering both the work time and finances necessary for frame development and maintenance. This is another area where survey training opportunities seem dispro-portionately low compared with the impact on surveys.

The (Revised) Survey Pyramid
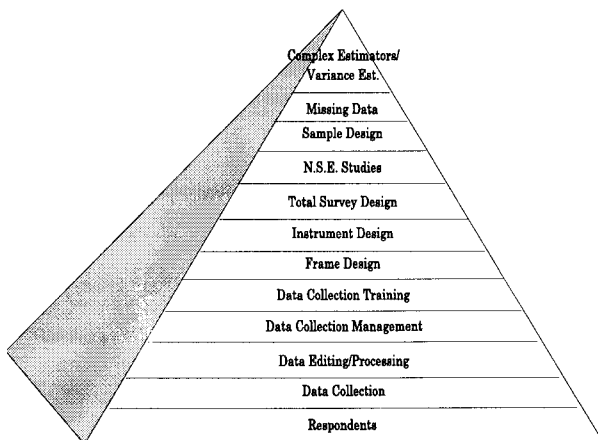
The first paper by Chapman, Moriarity and Sommers deals with the issue of complex association rules in an establishment survey frame. The authors do an excellent job of presenting the positive and negative aspects of selection alternatives aimed at burden reduction with multiple establishment firms (MEF's). My main suggestion for this paper is to define burden. The number of contacts or the frequency of contacts may not be positively correlated with burden. For example, would a large MEF retrieve the data from computers making it as burdening to report for one establishment as for many? See Cox, et al for more discussion. A second suggestion is to do a screening survey to find the small number of very large MEF's. A limited survey could provide valuable survey design information on size, retrieval method, retrieval location, and unioniza-tion, or other indicators of high intraclass correlation.

The paper contained limited data about the frames so I have some questions of a speculative nature. First, will domain estimates be made by state for large and small firms? If so, one might consider deeper stratification. If firms in establishments can be identified by state, consider some sample coordination method, possibly with permanent random numbers with different intervals in each state.

The intraclass correlation of .04 in the example seems low given the comment that it is reasonable that establishments in the same firm will have similar health insurance characteristics. Finally, as an aid to readers new to this area, Chaudhuri and Vos is a good reference to review when considering sample selection methods.

Edwards, Marker and Bond provide some data related to the frame discussed above. The paper is very useful because having some measure of the quality of the information used for survey design and management is important. My first discussion point is a note that the data presented in the paper are unweighted. I am left to assume that the design was not complex enough for the weights to matter.

My second point concerns the histograms used to display the data in figures 3 and 4 of the paper. Using the class breaks of 2, 1/2, 3, 1/3 . . . (of different size to the right and left of zero), with data that is typic-ally very skewed, may be misleading. A more inform-ative presentation might be a plot of log(survey data) versus log(frame data). Plot characters could be used to identify industry or other important factors and differ-ences in the ratio by size or clustering may be found if present.

The next question is directed to authors of the first two papers. Why is there no matched sampling from the 1994 survey in the 1997 selection? Matched samples could be very useful especially in the formation of policy analyses.

In the third paper, Steve Woodruff takes us into the difficult arena of estimating births and deaths in an establishment frame. The paper makes an excellent point, that finding a model that "estimates away" the problem of frame birth and death is not always possible. With many unpredictable shocks to the economy we may be unable to fit a model that will be satisfactory, so we are left with making a choice that "suffices."

The first comment on this paper concerns the composite estimate made from the direct and modeled estimates. If the direct estimate are not trustworthy, I worry about the variability of the estimated variance of the direct estimate being too influential in the weighting in the composite. A possible solution is to use a fixed value for the variance of the direct estimate based on historic data. As a final comment on this paper, I won-der whether one could use a double sampling scheme, first collecting information related to expectations for future increases or decreases in business activity, as a method to adjust for the shocks.

The paper by Byczkowski and Levy provides an excellent reference list for the many journal articles written on "multiplicity." This is an important contri-bution, since few sampling texts discuss frame con-struction. Even those texts with some discussion of frame issues may have little to say about many-to-many association rules. As one example, Dalenius describes the many-to-many association of sample units to pop-ulation elements as a difficult problem, but he provides little information concerning treatment. This may be the result of there being so few problems having sufficient structure to allow a workable solution. In this regard, I enjoyed this paper as a textbook example of sufficiently good auxiliary information to devise a workable many-to-many application.

I have two questions concerning this paper. First, what about nonresponse? A chill runs through me when I think about how difficult any nonresponse ad-justments would be in this context. The authors may be lucky in that it seems likely that nonresponse would be small in their example. Secondly, while a variance is developed, some comments on the difficulties one would face in trying to create an estimate of the variance would be worthwhile.

In the final paper of the session, Walsh and Clark gave a wonderful overview and case study of the extensive effort needed to maintain large complex frames. Particularly important points were:

- Frame staffs need to communicate what works with design and automation teams, and

- The impact of frame changes should be measured using parallel designs over time.

Not being a user of these frames, I cannot comment on the quality, customer service and accuracy results mentioned in the paper, but I do have one efficiency question.

I was not aware of the rural area frame activ-ities conducted by the Bureau. Since boundaries seperating the city and rural in Bureau work are likely to be close to the boundaries used in agricultural surveys, I wonder how much duplicative effort could be reduced with joint frame efforts. Several quality factors could benefit. Besides reduced costs, the potential exists for increases in information for design purposes due to heavier sampling and increased information sources. Further, frame revisions might be more timely.

As a concluding remark, maybe its time to start collecting well written and informative frame papers such as these into a collection to serve as a reference and material for sampling frame courses.

References:

Chaudhuri, A., and Vos, J. W. E., (1988), *Unified Theory and Strategies of Survey Sampling*, New Holland.

Cox, B. G., Binder, D. A., Chinnappa, B. N.,Christianson, A., Colledge, M. J., and Kott, P. S. (1995), *Business Survey Methods*, John Wiley and Sons, New York.

Dalenius, Tore, (1985), *Elements of Survey Sampling*, Swedish Agency for Research Cooperation with Developing Countries.

Fecso, R. S., Kalsbeek, W. D., Lohr, S. L., Schaffer, R. L., Scheuren, F. J., and Stasny, E. A. (1996), "Teaching Survey Sampling," *American Statistician*, forthcoming.