Discussion

Dale Atkinson, National Agricultural Statistics, USDA
Research Division, 3251 Old Lee Hwy, Rm 305, Fairfax, Va. 22030-1504

I appreciate the opportunity to discuss these papers. I enjoyed reading them and learned a lot from them. While each is very different in terms of approach, all five of them deal with the sometimes ugly business of identifying and handling (or assessing the impact of) "bad" survey data. As such they all document nonsampling error studies.

The basic reasons for nonsampling error studies of survey data are the following:

1) to correct the current survey data (if possible).

2) to head off future data problems (i.e., to discover training deficiencies, instrument wording problems, etc.).

3) to assess the potential impact of nonsampling errors on survey results.

Each paper in this session presents an approach of tackling one or more of these objectives. I'll now address each one individually by summarizing what I considered the key points of the paper and make a few comments.

The Weir, Emery and Walker paper documents developmental work on EIA's GEAQS graphical editing system, which combines and builds on editing products used elsewhere. This system features aggregate to individual level drill-down and bubble-up capability to pinpoint problematic areas. It makes effective use of standard exploratory data analysis tools, while visually distinguishing actual and imputed data and data of high and low influence on aggregates to prioritize follow-up. It was developed through an iterative approach of customer feedback.

My reaction to this paper was, "Ain't technology wonderful -- I want one!" It's almost scary how much analytic power we have at our fingertips today as compared to even 10 years ago. GEAQS is a terrific editing product which was clearly developed the right way -- with customer feedback along the way.

There are several issues that arise with the introduction of any new editing or analytic product. For example, how is updating handled? Does the system allow direct modification to the data that go to summary? If so, is there a built-in tracking system to monitor the editing process? Does the introduction of such a system encourage over-editing and/or render future data series incomparable to those of the past? How generalizable is the system? Considerable time and resources are involved in developing analysis tools. Hopefully we can build systems with a modular enough design to achieve maximum benefit from our developmental resources through relatively simple translations to other survey applications.

While these are issues that need to be considered with any new editing or analysis system, I'm very impressed with EIA's GEAQS system and commend all those associated with its development on a job well done.

The Thompson and Sigman paper was especially interesting to me. It is a very thorough discussion of a topic that NASS is also currently grappling with -- methods for developing ratio edit tolerances in statistical editing. In fact this is a paper that my unit, which is currently exploring the potential for statistical editing in NASS' surveys, is using as a key reference.

The authors describe various methods for setting tolerance limits for ratios and discuss the use of power transformations to obtain symmetry in the distribution of individual sample ratios. They compare edit tolerances based on robust vs. resistant procedures of handling outliers for many different ratios based on indicated Type I and II errors. The best statistical method (resistant fences) is then compared empirically with the operational (nonprobability) approach used for generating tolerance limits based on "gaps" in ordered ratios.

This paper has wide applicability since much of the editing done in business surveys can be expressed in terms of ratio edits. The authors appropriately spend considerable time on the handling of outliers, a very real and troublesome aspect of any analysis of survey data. They very clearly discuss the issues involved in dealing with ratio edits and compare several possible approaches for setting tolerance limits. They also conclude, I think very appropriately, that statistical methods are not a replacement for subject matter specialists, but can serve as a starting point, especially for ratios for which subject matter expertise is lacking.

As a result of its thoroughness, this was a very lengthy paper. The last version I received was 45 pages (which

was a trimmed-down version of the original draft). I'll be very interested to see how the authors reduce this to a 6 page "Proceedings" article. I was told that using a 1 point font was not an option! Be that as it may, this is an excellent paper. I especially highly recommend the complete and unabridged version.

The Harris paper was a descriptive one, summarizing the various editing and imputation approaches used in NCHS's registration systems and provider-based sample surveys.

Detailed editing is characteristic of most of Center's systems, and imputation is widely used in the provider-based sample surveys. Some of these utilize weight adjustments for unit nonresponse. Electronic data collection (i.e., CATI and CAPI) is increasingly common, with one-third of the data systems using CATI. 'Hard' editing cost figures were rare, but estimates of 10-40% of total survey budgets were typical. Slightly more than one-third of the data systems provide for monitoring of analysts/clerks in their data editing procedures.

The following results of the survey of data systems at NCHS were of particular interest to me:

18 (of 24) monitor their automated editing systems, but only 3 formally evaluate their systems.

21 (of the 24) maintained an audit trail for some or all data editing transactions, but the effect of data editing has been analyzed for only 5.

The good news is that NCHS maintains the data needed for editing analysis on many of its systems; the bad news is that, in general, little has been done with it up to now. In the near future I think much more analysis will be done in evaluating the impact of editing, as agencies and companies re-engineer their survey processes. Documentation of the current procedures, such as that presented in this paper, is a logical first step in getting there. I applaud the author's effort at taking this first step and pulling this information together.

The Sun paper discusses the use of control charts to monitor the CATI process in Statistics Canada's surveys. Two alternative procedures for specifying the upper control line (UCL) are compared based on coverage probability (CP), the percentage of in-control processes not flagged as out-of-control, and detection power (DP), the percentage of out-of-control processes flagged as out-of-control. The candidate procedures were a Poisson-based one using the average number of errors per enumerator per monitoring session, and a binomial-based one using the probability of an error estimated from the percentage of cases in which at least one error was detected for a given enumerator during a session.

Normal approximations of standard errors were used in both cases, and simulation studies were conducted by generating data from the hypothesized distributions with various drift factors. The author concludes that in terms of CP and DP the binomial and Poisson procedures performed similarly, which would recommend the binomial because of its additional simplicity.

The paper also discusses the importance of process stability in using control charts, and the techniques used to test for stability. Break points are used to subdivide the process into stable systems for control charting.

This was a good, readable paper which did a fine job of explaining the options and laying-out the issues. However, since there seems to be substantial real data available to work with, I would like to have seen some empirical results to back up the simulation results for the binomial and Poisson methods. Also, I would like to have seen some comparisons between these procedures and the operational (MBX) one. Selling either test procedure for operational use will require a demonstration that it works better than what is currently done. The draft paper I received doesn't address this type of comparison.

Finally, the McMahon paper documents a study to assess the impact on estimates of nonsampling errors in the clerical editing process, through analyzing a subsample designed for quality monitoring. Partnership returns were subsampled at a basic 10 percent rate with additional, nonrandom (normally larger than average) returns added later on. Estimates of nonsampling error (variability in the results of the two clerical edits) were calculated based on:

$$est.\ nse = \frac{\sum w_s * (clerk\ 1 - clerk\ 2)}{\sum w_s * (clerk\ 1)}$$

This study is a good example of trying to "mine" existing data for all the information we can. Unfortunately, as often happens where we do this (and get involved with data after the collection), some nonrandomness took place in the collection that's difficult (or impossible) to handle in analysis and estimation.

But, I guess resiliency in handling such situations is why we as statisticians get paid the "big bucks!!" Thanks again for the opportunity to discuss these papers. This was a very interesting and informative session.