

PROBLEMS OF MEASUREMENT IN SURVEYS  
DISCUSSION

Judith T. Lessler. NCHS

The papers in this session and those that are being presented at other sessions at this year's meeting represent a variety of fresh attempts to address problems of nonsampling errors in surveys. The current work has its roots in work done in the past. Survey researchers have studied a variety of sources of survey error through the use of data from independent sources, reinterviews, replicated or interpenetrated sampling, split ballot techniques, and double sampling/record check methods. These studies have led to improvements in survey design and practice. However, we face severe operational and financial difficulties in using these techniques to study nonsampling errors. Another problem that we have faced in studying nonsampling errors is the problem of decomposing and identifying the many, many factors that contribute to errors in surveys.

Many new attacks are being made against these problems. These include the use of latent structure models from the social sciences in which the concept of a latent variable that is not directly observable is used in the study of measurement errors. This is illustrated by the Tucker paper and the Cowen paper. Other work that is going on includes the investigations by O'Muircheartaigh at the US Census on relation of the simple response variance to various characteristics of the survey interview; the paper by Lavange and Folsom at a competing session on the regression estimates of survey operations effects to adjust for nonsampling errors. The paper by Pearl and Fairley on the use of strength of opinion to test for nonresponse bias is related to work being done by Swets and Salter at Bolt Beranek and Newman on the application of signal detection theory for studying reporting error. The paper by Mathiowetz is related to the new work that is going on concerning the use of cognitive theory to the study of reporting in surveys.

There are many other instances that could be cited, and I look forward to future ASA meetings and the tracking of these new efforts. Now I want to turn to the specific papers.

This session consists of 6 papers. Five of the papers are devoted to problems of measurement errors in

surveys -- that is, survey errors. One of the six papers has landed in the session because it is a survey of errors that lawyers make.

The papers on survey errors reflect a wide variety of approaches to the problems of measurement in surveys. We have two papers concerned with using statistical models and information collected within the survey itself to identify nonsampling errors. One paper is concerned with estimation in the presence of misclassified data under various models for the process that generated the errors. Two of our papers employ the more traditional approach of examining the agreement respondent reports and other sources of the same information. These papers go beyond the mere demonstration of differences and attempt to examine the processes that are generating the errors.

In spite of their diversity, I think that the papers make an interesting set because they are illustrative of the many new approaches and innovative techniques that are being developed for the study of survey errors.

The paper by Tucker represents the use of a technique that has a long history in the social sciences -- the study of latent variables. It is an interesting new approach to the study of measurement error. The model that is used assumes the reporting error is a property of the respondent related to his motivation to fill out the diary accurately and completely. Although this is a fairly reasonable assumption for a diary survey, it cannot be assumed to apply to the general survey situation where problems of recall, comprehension, self-presentation, interviewer actions, and so on are present as well as problems of motivation. Even within a diary survey, other sources of error can enter -- such as problems by the respondent in understanding what is to be recorded in the diary and how it is to be done, literacy problems or errors related to the skills required to complete the diary task, and the possibility that the diary modifies behavior. This last is particularly important given the use of the respondents description of his usual behaviour versus his actual behavior as a measure of quality.

The paper by Mathiowetz looks at the issue of recall loss and telescoping. Tables are presented which support the notions that recall is related to salience, length of recall period and number of distinct events to be recalled. This is an interesting paper in that it aims to investigate the reasons for failure to report spells of unemployment; however, given the extremely high level of underreporting, I feel that something other than lack of recall is going on.

Forty percent of the people who had been unemployed more than 5 weeks at a stretch during the past year failed to report such. The National Health Interview Survey asks people to report how many times during the past year that they stayed in bed for a half-day or more because of illness or injury. If 40 percent of people cannot recall 5 continuous weeks of unemployment, it is hopeless to ask a question such as that in the National Health Interview Survey.

I think that there is something other than recall loss that is operating. One possibility is the complexity of the question. The question was "Were there any periods since the beginning of the year before last, January 1981, when you were unemployed and looking for work or temporarily laid off for a week or more?" To answer "Yes", a person could simply have been laid off (presumably with an expectation of returning to work in the not too distant future) or he had to be both unemployed and looking for work. It is possible that the respondents felt that they were not to answer "Yes" unless they looking for work. Thus, as well as failure to recall, I think other explanations of the patterns seen in the data should be investigated.

The paper by Cowen uses an iterative technique, the EM algorithm,

to attempt to solve for the underlying true distribution given data classified by two faulty classifying techniques. It is interesting in that it presents the different assumptions that must be made in order to arrive at a solution. Also of interest is the use of the results to guide efficient double sampling schemes in which the true classification is obtained for certain cells. Noting that in some cases the algorithm converged to the correct solution if starting points that were near the correct values were used, I wondered if this fact could be used to guide the solutions.

The paper by Myers and Bernstein replicates a previous study on the relationship between respondent's selection of their job category and that of trained occupational coders. They note that some of the differences that are found are due to the fact that the respondent and the coders are not given the same conceptual task. The respondent is asked to choose the category that his job falls in from only a general description of the category. The occupational coders are given strict rules for classifying jobs depending on the description of the activities and occupation. Thus, the respondent and the coder have a different conceptual task. In some cases, the respondent could be expected to do a better job than the coders because they are more aware of the exact duties that the worker performs.

The paper by Pearl and Fairley is an interesting example of using information gathered in a survey and statistical models to assess the quality of the data. The notion that they develop is related to the idea of response probabilities, and while they are not directly able to measure the response probabilities and adjust the estimates for these differences, their procedure does allow for the detection of the potential for bias.