Journal of Official Statistics, Vol. 17, No. 1, 2001, pp. 51–54

Comment

Robert M. Groves[1] *and Nancy A. Mathiowetz*[2]

It is often the case that a bit of introspection provides a stimulus for progress in a field. We hope that the Platek and Särndal discussion might do so for survey methodology, but we believe the discussion might achieve that goal with a somewhat expanded viewpoint.

Most of the Platek and Särndal comments have a long history — the incomplete status of total survey error models, the multidisciplinary nature of official statistics, and design-based versus model-based inference. The tone of lament that pervades the writing, however, seems to center around the lack of progress on statistical models of total survey error.

A quick look at the list of references, however, reveals that the article does not review the development of theory that underlies important measurement innovations in surveys in the last 15 years. In many of these discoveries, the ingredients Platek and Särndal seek are present – well-elaborated theory exists; the theory informs survey practice; and there is replicated scientific evidence that the practice leads to higher survey quality. For example, the article fails to reference the discovery, now well replicated, of consistent reduction of social desirability bias for threatening questions using self-administration (see Tourangeau and Smith 1996; De Leeuw 1992; even Nathan et al. 1990). Cognitive psychological theories provide insight into how people reconstruct the date of an event. This framework permitted the reduction in dating errors through the use of relative dating as opposed to absolute dating of events (Loftus and Marburger 1983). For other measurement goals, the theory led to reduced dating errors through the use of bounded recall techniques (Neter and Waksberg 1964) and two time-frame methods (Sudman, Finn, and Lannom 1984). Cognitive psychology also provides the theoretical basis for understanding the influence of behavioral complexity on retrieval strategies used in answering behavioral frequency questions (Blair and Burton 1987; Burton and Blair 1991; Sudman and Schwarz 1989). Questionnaire designers, informed by these findings, direct respondents toward retrieval strategies specific to the analytic goals of the question. The field has derived the theoretical basis for understanding context effects in attitude measurement (Sudman, Bradburn, and Schwarz 1996). This had led to practical guidance regarding the placement of general attitude items prior to specific attitude items. Theories concerning psychological norms of reciprocity underlie the consistent performance of incentives, especially the superior performance of prepaid incentives over promised incentives. New theories concerning psychological tradeoffs in decisions explain how incentives can affect nonresponse error (Singer, forthcoming).

[1] University of Michigan and Joint Program in Survey Methodology, College Park, MD, U.S.A.
[2] University of Maryland and Joint Program in Survey Methodology, College Park, MD, U.S.A.

There are three features these developments have in common:

1. The theories they employ are theories of human behavior or cognition, not theories of probability and statistics.
2. The theories lead to methods that produce lower statistical errors (generally response bias or response variance).
3. The success of the methods is demonstrated in replicated randomized experiments in a variety of survey settings. However, the theories themselves are not stated in terms of statistical error models.

We agree with Platek and Särndal that total survey error models, as models of the statistical properties of survey statistics, have not evolved over the last 15-20 years to be more useful to the practitioner. We disagree with one apparent implication of that viewpoint, however. We do not agree that state-of-the-art practice is using the same theories to design surveys as was true 15-20 years ago. Further, we believe that the behavioral and cognitive theories underlying practice in measurement design, for example, are as useful and robust as many of the statistical theories guiding sample design and estimation.

Given this, let us parse the sentence, ''Can a statistician deliver?'' First, Platek and Särndal clearly use a broad definition of ''statistician,'' including ''theoretical statistician, survey methodologist, subject matter specialist, information technologist, and survey manager.'' We agree. We believe such a definition is appropriate because, as the theoretical and practical developments in surveys over the past 10 years attest, the field progresses effectively only with a smooth interplay of the multiple disciplines that constitute it. With regard to ''deliver,'' it seems the authors operationalize this concept with two key questions: ''Is the statistician capable of delivering not only numbers but also adequate quality assurance?'' and ''What forms should the assurance take?''

Do we deliver adequate quality assurance? Are we getting better? It seems clear that we have tools to improve quality that were absent some time ago. The tools have theoretical grounding (but again, increasingly theories of human cognition and behavior). The empirical support for the theories comes from the preponderance of evidence from methodological investigations in a wide variety of survey circumstances. The theories are often tested using randomized experimental designs imbedded in a survey.

Platek and Särndal ask, ''What forms should the assurance take?'' What is different between evidence of ''quality'' from the cognitive and behavioral theories and those offered by statistical is that the researcher must logically argue that the essential survey conditions (using the term in the Hansen, Hurwitz, and Bershad sense) of his or her survey are equivalent to those of the various replications of the theoretical result. Statistical theories yield estimates of error properties as part of their practice; the behavioral and cognitive theories do not automatically do so.

Since most of the error models are off-shoots of variance components models, however, they often offer little practical guidance to the survey researcher concerned about reducing error. Take, for example, the simple and powerful model of Hansen, Hurwitz, and Bershad (cited by Platek and Särndal) regarding the interviewer component of variance for a sample mean. With the formulation that the inflation of variance is a function of workload size (m) and an intraclass correlation (ρ) expressed as $[1 + \rho(m - 1)]$, the practitioner is pointed in the direction of reducing interviewer workload as a way to reduce interviewer

variance. However, because the model is based only on variance components and does not have a parameterization that identifies the *causes* of the variance components, use of the model as a practical tool is unwise. One example illustrates this. The model assumes that the intraclass correlation is independent of workload size, yet most statisticians who design and manage field interviewing staffs know that hiring more interviewers generally leads to increased interviewer variability (mainly because training protocols do not provide guidance for all field situations interviewers encounter). Hence, reducing workload by hiring new interviewers often increases ρ. Only in the restricted case of drawing more interviewers from precisely the same population of interviewers does the practice yield the desired result. The general point may be that total survey error models, as mere expressions of the magnitude of various error components, are not useful to the survey designer. If they were reparameterized as models of components of various causes of the statistical errors, then designers could be directed to ways of reducing the errors.

The advantage of the behavioral and cognitive theories of response error is that they tend to identify the causes of the errors. Through that they often give clear guidance on the reduction of error. The disadvantage of the same theories is that they have not been translated into models of statistical error. Because of that they do not provide the statistician with automatic measures of the error reduction. While we agree with the perceived stagnation in the evolution of total survey error models, we have hope that the integration of the behavioral theories with statistical models may give the field new energy.

Can the statistician ''deliver?'' We are getting better at it with each passing year, but we are currently not providing evidence of this for each statistic we deliver. To do that, the powerful behavioral theories now benefiting survey practice need to be translated into statistical models of error.

References

Blair, E. and Burton, S. (1987). Cognitive Processes Used by Survey Respondents to Answer Behavioral Frequency Questions. Journal of Consumer Research, 14, 280-288.

Brown, N.R., Rips, L.J., and Schevell, S.K. (1950). The Subjective Dates of Natural Events in Very-Long-Term Memory. Cognitive Psychology, 17, 139-177.

Burton, S. and Blair, E. (1991). Task Conditions, Response Formulation Processes, and Response Accuracy for Behavioral Frequency Questions in Surveys. Public Opinion Quarterly, 55, 50-79.

De Leeuw, E.D. (1992). Data Quality in Mail, Telephone, and Face-to-Face Surveys. Amsterdam:T.T.-publikaties.

Hansen, M.H., Hurwitz, W.N., and Bershad, M.A. (1961). Measurement Errors in Censuses and Surveys. Bulletin of the International Statistical Institute, 38, 351-374.

Loftus, E. and Marburger, W. (1983). Since the Eruption of Mt. St. Helens, Has Anyone Beaten You Up? Improving the Accuracy of Retrospective Reports with Landmark Events. Memory and Cognition, 11, 114-120.

Nathan, G., Sirken, M., Willis, G., and Esposito, J. (1990). Laboratory Experiments on the Cognitive Aspects of Sensitive Questions. Paper presented at the International Conference on Measurement Error in Surveys, Tucson, U.S.A., November.

Neter, J. and Waksberg, J. (1964). A Study of Response Errors in Expenditure Data from Household Interviews. Journal of the American Statistical Association, 59, 18-55.

Singer, E., (forthcoming). Incentives in Survey Research. In Survey Nonresponse, eds. R. Groves et al., New York: Wiley and Sons.

Sudman, S., Bradburn, N., and Schwarz, N. (1996). Thinking about Answers: The Application of Cognitive Processes to Survey Methodology. San Francisco: Jossey-Bass.

Sudman, S. and Schwarz, N. (1989). Contributions of Cognitive Psychology to Advertising Research. Journal of Advertising Research, 29, 43-53.

Sudman, S., Finn, A., and Lannom, L. (1984). The Use of Bounded Recall Procedures in Single Interviews. Public Opinion Quarterly, 48, 520-524.

Tourangeau, R. and Smith, T. (1996). Asking Sensitive Questions: The Impact of Data Collection Mode, Question Format, and Question Context. Public Opinion Quarterly, 60, 275-304.

Received December 2000