Journal of Official Statistics, Vol. 17, No. 1, 2001, pp. 21–24

Comment

Barbara A. Bailar[1]

The title of the article by Platek and Särndal inspires curiosity but I soon realized that it was confined to survey research methodology. The title does not fit even this narrower topic, especially since there are many disciplines involved in survey methodology.

The overall tone is pessimistic, as expressed in the statement, ''...we still have a weak theoretical basis for the production of official statistics.'' This is true only if one believes that statistics should be a strictly ''theoretical'' discipline in the narrow mathematical sense, and it denies the benefits of organized experience in understanding data. Not all theory is mathematical. (Chemistry comes to mind.) Examples of conceptualizations and generalizations common in national statistical agencies include a vast body of knowledge about how to ask questions, issues of privacy and confidentiality, techniques for coping with nonresponse, and editing of files.

When I first began my career, I was fortunate to work at the U.S. Census Bureau when the legendary statisticians were still there; Morris Hansen, Bill Hurwitz, Max Bershad, Ben Tepping, and many others. One of the most exciting things about that environment was that practice was feeding the development of theory and the theory was informing practice. I read the papers these remarkable people wrote, and saw how, for example, sampling with probability proportionate to size went from a good idea that seemed to work to a technique based on mathematical theory. There were many advances along these lines. At staff meetings, someone would say, ''This seems to work. I have been working out why it works. Let me show you.'' It was an exciting place.

Hansen, Hurwitz, and others expanded their view to matters beyond sampling. Thus, the ''mean-squared error'' model that was the basis of the experiments about the effects of interviewers on survey responses came about. Within that same kind of model, they looked at biases caused by respondent errors in reporting. The U.S. Census Bureau's move to self-enumeration was based largely on these studies. (See Hansen and Marks 1958.)

In fact, one of the things I look back on with some amazement is how quickly Hansen and Hurwitz were able to get the results of experiments and models into practice. Things take longer now. So perhaps my attitude of optimism and the feeling that survey methodology has made enormous strides in the last 50 years is based on my fortunate circumstances of working for 30 years in a place where survey theory and methodology grew together, and intersected to produce great gains.

Platek and Särndal briefly discuss the use of Current Best Methods, or ''best practices'' as they are known in different settings, and lament the fact that the procedures used today may not be ''best'' tomorrow. Cannot other scientific fields be similarly cited? I think of how medicine, chemistry, and astronomy constantly change. There are very few, if any, absolutes.

[1] National Opinion Research Center, Chicago, IL, U.S.A.

When we add new and different ways of responding to surveys, there are going to be changes in the error structure of survey data.

Survey methodologists must be concerned with all aspects of surveys, not just design, or instruments, or field work, so we examine each piece of the survey. Even if we had a total survey error model that would express the accuracy of a survey result in one number, just as we have a sampling variance to reflect precision, we would still want to look at the individual parts of the survey. That is why there have been efforts to produce error profiles or quality profiles, such as for the Current Population Survey, the Survey of Income and Program Participation, and some agricultural surveys. The U.S. Census Bureau has recently made a commitment to produce quality profiles for all of its surveys.

We have learned a lot in the last fifty years. Questionnaire construction has received a great deal of attention over the last 15 years. When cognitive psychologists join statisticians to learn more about what people understand when they hear survey questions, the results are quite impressive. Equally important is the great progress in question wording, question placement, the full statement of alternatives in a question, and the meaning of questions in regional and international contexts. Overall, this work has greatly enhanced the quality of survey research. Much work has been done to measure the bias introduced inherent in different question wordings or placements. However, to reinforce the idea that there are no absolutes, an interesting controversy has arisen between those who want a standardized survey interview in which each interviewer asks the questions exactly as worded and those who promote conversational interviewing in which interviewers can tailor the questions to the respondents' needs or understanding. We will not know the outcome of this controversy for several years, but the results will probably change the way survey researchers frame questions.

Another area in which theory has improved practice is imputation for item nonresponse. Survey practitioners now have available a wide range of options. When imputations were first made for census data, the information for the same person was repeated multiple times. For establishment surveys, average amounts were imputed multiple times. Gradually, practice became based on the high correlation of near neighbors and theory followed. Now, there are systems of multiple imputation from which survey practitioners can calculate the variance in survey statistics caused by imputation. The bias and increased variance caused by imputation, contrary to the authors' report, are discussed frequently. Record-check studies have evaluated different imputation methods and their effects on different items. Much more is known now than twenty years ago. Some surveys use multiple imputation and report imputation variances. (See Kennickell 2000.)

Usage has improved considerably since the days when the use of averages for imputations caused estimated variances to be biased downward.

Recent theoretical work has also increased knowledge and improved practice in the realm of coverage. Research on coverage, at least in the large statistical agencies, became more intense as a result of the concern about census coverage, especially the undercoverage of minority groups. (See Bailar 2000.)

Many more studies were conducted to measure coverage. The Current Population Survey has for years, estimated coverage and tried to remedy problems by poststratification. In income surveys, there are coverage errors at both ends of the income spectrum. Telephone surveys omit the low income groups. The recent emphasis on studies of individuals

on welfare, applying for welfare or newly off welfare, brought to light many of the problems of contacting low income populations. Similarly, many surveys omit those people who have no usual place of residence or live on the street. For many survey topics, this can no longer be done. Also, many surveys regularly omitted respondents who did not speak English or Spanish. Because of high immigration and the importance of immigrants to our understanding of economics, health, and education, the immigrants must now be included. We now have tools to measure the bias caused by coverage error, including record checks and reinterview surveys. Learning how to avoid the problem is one in which practice and ingenuity intersect with theory to lead the way.

Total survey error models are found lacking by the authors. Again, is the glass half full or half empty? The remarkable gains in understanding interviewer variability came because of the Census mean-squared error model. Record check studies gave us insights about many survey items, such as finding that wage and salary income are reported rather accurately, but other kinds of income are reported less well. Using some of the estimates of parameters in the Census survey model has led to distinct improvements in data. For example, what do we learn when the simple response variance is computed after a reinterview with respondents, and again when a coder recodes a subsample of cases, and we find the index of inconsistency is high in each case? My reaction is to look at the question, because the ambiguity resides not only in the respondents, but also in the coders. Is this theory? It is not mathematical theory, but it can be put in a list of *''do's and don't's''* for theory in survey research just as there are such lists of *''do's and don't's''* for clinical trials.

In conclusion, I review the five points made by the authors:

1. *Quality is multidimensional.*
   Because this is true, survey research attracts people of many different skills and from many different disciplines. Statistics is only one important discipline involved in the enterprise, but what makes survey research so fascinating is the interweaving of all the disciplines. Quality in survey design does not assure quality in questionnaire conceptualization, or in the psychology of motivating interviewers, or in coverage. Statisticians among many others, deliver time and time again when we improve the quality of our instruments, the understanding of interviewers, the underpinnings of imputation and we explain these things in language understandable to users of data.
2. *Survey methodology is a set of practices with no unifying theory covering the entire process*. This of course refers to no unifying mathematical theory. Statistical agencies have a great deal of knowledge at a sufficiently abstract level that they can bring to bear on every survey. Practitioners have built up a codification of practices over the years and many of those serve as a unifying mechanism. In many ways, this is no different from other areas of application in which statistics is one among many disciplines: clinical trials, weather forecasting, and population estimation, to name a few.
3. *The user wants confidence in both survey design and survey process.*
   I disagree with the authors that the users cannot or prefer not to question accuracy. Many users do question accuracy. Census results have been changed because users

found an error in the coding of income. Users have worked with statistical agencies in formulating questions so that survey methodologists will understand basic concepts. Users bring alternative data sets to the fore for agencies to explain differences. Users care much more about accuracy than about just timing and coherency. Users do not want to make policy errors because of faulty data. The difference between my point of view and that of the authors may reflect differences among user communities. I have often dealt with very sophisticated users who wanted to know about the effects of imputation, why question wording made a difference, what the estimated coverage of the survey was, why there were inconsistencies in public use files, etc. These kinds of users are often found in Congressional staff, among academic users, and among business users. The casual reader of a news story may, of course, be different.

4. *Statistical theoreticians have focused on sampling variance and neglected bias.* I agree whole-heartedly and a substantial body of theory about bias has come from other disciplines. It is to the loss of the survey methodology world that statisticians have put so much emphasis on relatively small gains in variance reduction while there are large gains to be made in reducing bias. However, other disciplines have stepped in and added value.
5. *Total survey error does not yet provide a systematic evaluation of total accuracy.* However, there has been admirable and substantial progress at looking at many of the specific error components. We know a lot more than we did 50 or even 20 years ago. The glass is filling steadily and is more than half full.

Platek and Särndal might argue that none of these advances are ''theoretical'' and that hence such advances do not undercut their argument, but that argument relies on a narrow definition of theory. A broader view shows that theory has greatly enhanced every aspect of survey work, and there is good reason to think that productive theoretical results will continue for a long time.

References

Hanson, R. H. and Marks, E.S. (1958). The Influence of the Interviewer on the Accuracy of Survey Results. Journal of the American Statistical Association, 53, 639-655.

Kennickell, A.B. (2000). Wealth Measurement in the Survey of Consumer Finances: Methodology and Directions for Future Research. Paper prepared for the 2000 annual meeting of the American Association for Public Opinion Research.

Bailar, B.A. (2000). The Past is Prologue. Survey Methodology, 26, 21-30.

Received December 2000