

## Comment

*Stephen E. Fienberg<sup>1</sup> and Judith M. Tanur<sup>2</sup>*

Don Dillman has written a thoughtful and provocative article on the processes surrounding innovation in government surveys and in the course of doing so has shared with us insights about organizational change more broadly. We find ourselves in broad agreement with his diagnosis of difficulties, especially as they relate to the problems posed by different cultures in the research and operational parts of organizations. Further, we note that government agencies are particularly loath to make changes in practices and procedures because they often have a public responsibility to maintain continuity of their data. We also share with him a strong interest in the study of measurement and non-response errors and in understanding their implications for survey design. Where we differ with him, however, is in matters of the history of the development of government surveys, in some of the details of the response of U.S. government agencies to technical innovations over the past decade, and in some of the agencies' own initiatives in the area of measurement of error.

Dillman's observations grow out of his experiences at the U.S. Bureau of the Census from 1991 to 1995, during which time he worked primarily on the redesign of the decennial census. The U.S. census poses very special obstacles to innovation and thus generalizing from it to other government surveys is particularly hazardous. The decennial census appears stuck in the rut of a ten-year cycle that inhibits major change. Enthusiasm for research ideas early in the decade is severely tempered by the operations staff who insist on large scale tests in circumstances resembling an actual census, and rightfully so. The mid-decade test censuses thus offer the only real opportunities for mounting carefully controlled experiments and creative research ideas are often discarded before they can move towards serious implementation. But the agents of conservatism are just as often the statisticians as the operations staff. One of us (Fienberg) has tried for years without success to convince statisticians at the bureau to test a way to assess heterogeneity and dependence in the survey methodology used to measure the differential undercount.

While he spends much of his article discussing non-sampling errors in surveys, Dillman fails to note that the recent focus of research statisticians in government statistical agencies on measurement and non-sampling errors is part of a cyclical pattern

<sup>1</sup> Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213-3890, U.S.A.

<sup>2</sup> Department of Sociology, State University of New York at Stony Brook, Stony Brook, NY 11794-4356, U.S.A.

going back to the 1940s (e.g., see Fienberg and Tanur 1990 and the references therein). We believe that it is important for readers to appreciate the extent over interest in this topic among the research statisticians at the U.S. Bureau of the Census of the past 50 years. Bureau statisticians helped to pioneer the notion of models for the integration of sampling and non-sampling errors in the early 1950s and these have served as exemplars for those in the field even to the present day. Thus the models in Groves (1989), for example, can be viewed as elaborations of those introduced by Morris Hansen and his collaborators (e.g., see Hansen et al. 1951). The thinking that lay behind that model was in fact responsible for the shift, begun experimentally in 1960 and implemented in 1970, from a personal interviewer based census to one involving mail-out-mail-back self-administered census forms. This shift eliminated several components of correlated response error due to interviewers and was supported by a carefully controlled experiment, carried out as part of the 1960 decennial census. Eighteen years later, the preparation by Camilla Brooks and Barbara Bailor of a landmark review of sources of error in the Current Population Survey (CPS) (Subcommittee on Nonsampling Errors 1978) signaled a new era in survey design that led in part to the recommendations of the National Commission on Employment and Unemployment Statistics (1979) and ultimately to the redesign of the CPS from 1989–1993 that was fully implemented in January of 1994 (e.g., see the description in Norwood and Tanur 1994).

Dillman suggests that it is the research statisticians who are concerned with sampling error and that non-sampling errors are the province of those in operations – and that those concerned with operations give short shrift to issues of measurement error. The movement to develop the Cognitive Aspects of Survey Methodology (CASM), initiated over a decade ago (e.g., see Jabine et al. 1984, as well as the updates by Tanur 1992, and Tanur and Fienberg 1992), was intended to break from such a stereotype and to bring statisticians, both in government and out, back to the focus on non-sampling error. The extent to which CASM has been successful can be measured in various ways, but we draw attention, in particular, to the three government agency cognitive laboratories in the National Center for Health Statistics, the Bureau of Labor Statistics (BLS) and the Bureau of the Census. The interdisciplinary nature of these laboratories and the effect that they have had on the design of surveys across the government has been dramatic. This does not mean that Dillman's concerns are unwarranted, but rather that he may be attempting to generalize inappropriately from the very special circumstances of the decennial census to other government survey enterprises.

The recent joint Census-BLS effort to redesign the questionnaire of the CPS concurrently with the decennial revamping of the sample design seems to us a shining example of the innovation possible in government statistical agencies. In a multi-year effort, teams composed of personnel from the two agencies pinpointed areas of the questionnaire thought to be troublesome for respondents or to yield less-than-ideally valid data. Using a variety of techniques from various parts of the social sciences, including focus groups, behavior coding, card sorts, and debriefing probes, researchers sought to establish whether these questions were indeed problematic and then they tested various ways of solving the identified problems.

After several iterations, the research produced a new questionnaire that, amongst its other features,

- is fully automated, using CAPI and CATI.
- uses dependent interviewing for selected questions. (For example, the interviewer now ascertains occupation and industry by telling respondents how their answers were recorded in the previous interview and then asking whether there has been any change – this appears to eliminate a good deal of artifactual change in these variables.)
- takes seriously respondents' definitions of terms. (For example, the CPS's meaning of "on layoff" entails the expectation of recall to a job within a short period of time; the research showed that respondents use the term as a euphemism for "fired". The CPS now explicitly asks about the expectation of recall.)
- alters the opening survey question about activities in the previous week to include explicitly work in a family business and to remove a subtle but important bias that encouraged the classification of many women as not in the labor force. This change alone appears to have accounted for most of the change in the measured unemployment rate (about 0.4%) that coincided with the introduction of the new questionnaire.

How is it possible to attribute change in the measured unemployment rate to questionnaire change? Not only was there careful testing a priori, but for 18 months the Census Bureau conducted the old and the new surveys in tandem so that statisticians could calibrate the changes attributable to the altered methodology and then tease these apart from those traceable to changes in the underlying phenomenon under measurement. Yes, the changes took a carefully orchestrated program of research carried out between the two agencies over several years, but the planning for change also respected the need for continuity in measurement. Thus, when the Census Bureau and BLS introduced the altered design and questionnaire with briefings for economic reporters and members of congressional oversight committees, the changes hardly created a ripple. This is an ideal way for statistical agencies to accomplish fundamental change.

Dillman rightly spends much time focusing on the problems associated with organizational hierarchies, and he has also correctly identified technology as one mechanism for breaking down barriers and bringing about institutional change. Social research has convincingly demonstrated how the introduction of e-mail from top to bottom in an organization lets workers communicate and interact with administrators with surprisingly few inhibitions (Sproul and Kiesler 1991) thus diminishing the verticality of the organization. Dillman suggests that government survey organizations face tasks unlike those usually faced in universities, but large organizations often have many levels of hierarchy. One of us (Fienberg) had recent experience as a senior academic administrator in one of Canada's largest universities. There was certainly a large operational bureaucracy at work, and on many occasions dealing with it was enormously frustrating. But the worst bureaucracy was on the academic side, and the researchers were reluctant to give up their elaborate mechanisms of perceived control, at the very same time as the operations people were embracing

the changes associated with new computer technology intended to help both parts of the organization. In fact, those on the operations side often originated creative ideas for changes on the academic side of the university.

As a senior research fellow at BLS, the other of us (Tanur) had the opportunity a few years ago to observe a federal statistical agency from the inside. That experience included sitting in on some of the meetings of the Census-BLS interagency task force redesigning the CPS and interacting with agency staff from both operations and research. Input from representatives of both cultures in the planning for innovation strikes us differently than it does Dillman; we see people from operations often contributing valuable insights into what has been found to work in the past and what needs improvement. There is always the danger that such contributions will be a force for maintaining the status quo, or at least for minimal change, but if we follow Dillman's analogy, were pilots and flight attendants to be part of the design for a new airplane, the outcome might well be a safe and fuel efficient airplane that is easy to handle, easy to enter and exit, and that serves tasty food.

## References

- Fienberg, S.E. and Tanur, J.M. (1990). A Historical Perspective on the Institutional Bases for Survey Research in the United States. *Survey Methodology*, 16, 31–46.
- Groves, R.M. (1989). *Survey Errors and Survey Costs*. New York: John Wiley.
- Hansen, M.H., Hurwitz, W.N., Marks, E.S., and Mauldin, W.P. (1951). Response Errors in Surveys. *Journal of the American Statistical Association*, 46, 147–190.
- Jabine, T., Straf, M., Tanur, J.M., and Tourangeau, R. (eds.) (1984). *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines*. Committee on National Statistics. Washington, DC: National Academy Press.
- National Commission on Employment and Unemployment Statistics (1979). *Counting the Labor Force*. Washington, DC: U.S. Government Printing Office.
- Norwood, J.L. and Tanur, J.M. (1994). Measuring Unemployment in the Nineties. *Public Opinion Quarterly*, 58, 277–294.
- Sproul, L. and Kiesler, S., (eds.) (1991). *Connections: New Ways of Working in the Networked Organization*. Cambridge, MA: MIT Press.
- Subcommittee on Nonsampling Errors (1978). *An Error Profile: Employment as Measured by the Current Population Survey*. Prepared by C.A. Brooks and B.A. Bailar. Federal Committee on Statistical Methodology, Office of Federal Policy and Standards. Statistical Policy Working Paper No. 3. Washington, DC: U.S. Department of Commerce.
- Tanur, J.M. (ed.) (1992). *Questions About Questions: Inquiries into the Cognitive Bases of Surveys*. New York: Russell Sage Foundation.
- Tanur, J.M. and Fienberg, S.E. (1992). Cognitive Aspects of Surveys: Yesterday, Today, and Tomorrow. *Journal of Official Statistics*, 8, 5–17.