

# Notes on Nonsampling Errors

## for Statistics 36-303: Sampling, Surveys, and Society

Stephen E. Fienberg  
Department of Statistics  
Carnegie University

January 23, 2003

### I. Types of Errors

There are two types of errors associated with sample surveys: sampling error and non-sampling error. Many sampling statisticians devote their energies to the sampling designs associated with surveys and they use probability and statistical theory to develop designs with the smallest variances for certain estimated quantities. Others devote their efforts at minimizing nonsampling errors such as nonresponse and measurement errors, but they cannot rely on statistical theory to do so. Instead, they study actual human behavior and accumulate empirical evidence about different types of nonsampling error and changes in survey practice such as questionnaire design or interviewer training that might lessen their impact. The successful survey depends on a trade-off between these two different types of errors.

In a perfect survey world, where respondents all respond and answer questions truthfully and without error, all we need to worry about is sampling error. Statistical theory allows us to formulate estimators that are unbiased, i.e., correct on average over repetitions of the sample. or at least approximately so for large samples. Thus sampling error contributes only to the variability associated with our sample estimates.

In the less-than-perfect survey world, nonsampling errors occur and contribute to both bias and increased variability. Bias represents a form of systematic error that does not go away as the sample size increases and it occurs when, for example, those who respond are basically different on average from those who do not. But in addition nonsampling errors may have a non-systematic component and this gets added into the overall variability of the estimate that the survey produces. For example, respondents might answer questions differently if the questionnaire is administered repeatedly, but on average they give the correct response. Then the result is a form of response error that needs to be combined directly with the sampling error to get a revised variance for inference purposes.

One way to combine the two types of error, or more explicitly bias and variance, is by looking at what we have come to know as Mean Squared Error:

$$\begin{aligned}\text{Mean Squared Error} &= \text{Average Value of (Estimate - True Value)}^2 \\ &= \text{Variance} + (\text{Bias})^2.\end{aligned}$$

Unfortunately, in many surveys all we know is the extent of sampling variability and at best a report will remind us that estimates may also be subject to various nonsampling errors. We occasionally see such statements in newspaper accounts of surveys such as in the boxes describing margins of error in *The New York Times*. In most surveys there are no such warnings, but the problem of nonsampling error may be substantial.

## II. A Basic List of Errors

Amidst the flurry of activity on the theory and practice of probability sampling during the 1940s, attention was being focused on issues of nonresponse and other forms of nonsampling error. In a review of work on errors in surveys, W. Edwards Deming (1944) listed 13 factors affecting the ultimate usefulness of surveys:

1. variability in response;
2. differences between different kinds and degrees of canvass;
3. bias and variation arising from the interviewer;
4. bias of the auspices;
5. imperfections in the design of the questionnaire and tabulation plans;
6. changes that take place in the universe before tabulations are available;
7. bias arising from nonresponse (including omissions);
8. bias arising from late reports;
9. bias arising from choice of date for the survey, or of the period covered;
10. bias arising from an unrepresentative selection of respondents;
11. sampling errors and biases;
12. processing errors (coding, editing, calculating, tabulating, tallying, etc.);
13. errors in interpretation.

With the exception of sampling error in item 11, this is a list of nonsampling errors and most of these errors had been or would become the focus of research by statisticians at the Bureau of the Census, where Deming worked at the time he wrote this paper.

More modern terminology for errors in Deming's list would distinguish among frame errors, sampling errors, errors of non-observation (e.g., nonresponse error including coverage error), and observational errors (e.g., response errors due to interviewer, instruments, respondents, mode of data collection, recording, coding, etc., as well as measurement error due to inaccuracy of measurement or some other type of intrinsic error).

### III. Questionnaires, Interviews, and Cognition

Much attention has gone into the issue of biases and errors associated with questionnaire design and with the interview context. As a result of systematic efforts to improve survey methodology over the past forty years, survey researchers have evolved a highly developed art of questionnaire design and interview procedures to reduce nonsampling errors (e.g., see Payne, 1951), such as those described in Deming's list above, and they have carried out many scientific studies to test aspects of that art (e.g., see Schuman and Presser, 1981). Until recently, however, research on understanding the survey interview situation has been relatively unsystematic. The recent change came, in part, through the recognition that other fields, in particular cognitive psychology, had insights that would assist survey researchers examine the interview process.

Among non-sampling errors are those occasioned by the cognitive processes that respondents and interviewers are required to exercise in the survey interview situation. Respondents must often recall events and make judgments or estimates, and must always face issues of comprehension of the questions asked – their meaning to respondents as well as their meaning to interviewers. Survey researchers are now beginning to draw on the concepts of cognitive psychology and the expertise of cognitive psychologists to investigate more systematically these issues of non-sampling error.

Survey researchers now view cognitive aspects of survey design as a vital part of the survey enterprise, and government statistical agencies such as the Bureau of the Census, the Bureau of Labor Statistics, and the National Center of Health Statistics have established cognitive laboratories to assist the development of new questionnaires and the reexamination of old ones. Tanur (1991) reviews much of the progress in this area that has taken place over the past decade, and many of the recent changes in the questionnaire for the Current Population Survey drew on these developments (see Plewes, 1994).

### IV. The Evolution of a Survey Question

The following question has traditionally been asked by the National Center for Health Statistics (NCHS) and the Census Bureau as part of the National Health Interview Survey (NHIS),

which is administered to approximately 40,000 households annually in face-to-face interview:

How many times did you visit a doctor in the last two weeks?

The actual wording they have used has evolved in interesting ways over the years and it was never as simple as the question suggests (Statistical Policy Working Paper 10, 1983).

To understand the question wording you need to know that the question appears following questions on hospitalization. The NCHS interview goal is to get additional doctors visits and so when respondents have reported hospitalizations they are reminded not to include them in the followup question. In the 1970s, NCHS used to ask a household respondent using three probes about doctors visits:

1. During the past 2 weeks, how many times did anyone in the family see a medical doctor? (Do not count doctors seen while a patient is in the hospital).
2. During that 2-week period, did anyone in the family go to a doctor's office or clinic for shots, x-rays, tests or examinations?
3. During that period, did anyone in the family get any medical advice from a doctor over the telephone?

The statisticians at NCHS believed from interviewer and other information that the concept of physician visits was not being understood and that respondents may not be including specialists, including visits where they were seen by a medical assistant, or phone calls to obtain prescriptions, etc. Thus they devised some alternative questions for use in early 1979 on 260 households in a pilot test. They used the following lead-in to 4 questions:

The next questions determine whether anyone has recently received health care from any kind of medical doctor-including general practitioners and any type of specialists, such as pediatricians, psychiatrists, ophthalmologists, and so forth. Also include health care received from a doctor's assistant or a nurse working under a medical doctor's supervision.

They discovered that their new attempt was too wordy and respondents frequently interrupted the interviewers. This led to additional wording changes that were tested in a "split-sample" survey experiment conducted in the 4th quarter of 1979. The control group of 10,500 households received the standard questionnaire and experimental group of 5,000 households received the new version with the following altered questions about doctors visits.

These next questions are about health care anyone in the family may have received.

1. During the past 2 weeks [outlined in red on that calendar], how many times did - see or talk to a medical doctor? Include all types of medical specialists, such dermatologists, psychiatrists, and ophthalmologists, as well as general practitioners. (Do not count times while an overnight patient is in the hospital).
2. We are also interested in the number of times anyone received health care from a nurse or anyone else working with or for a medical doctor. (BESIDES THE TIME(S) YOU JUST TOLD ME ABOUT) During those 2 weeks did anyone in the family receive care at home or go to a doctor's office, clinic, or hospital to receive health care?
3. (BESIDES THE TIMES YOU HAVE ALREADY TOLD ME ABOUT) During those 2 weeks did anyone in the family get any medical advice over the PHONE from a doctor, nurse, or anyone else working with or for a medical doctor? Include calls to get prescription or test results.

The new version was a success, in that it produced the desired increased reporting patterns, but residual awkwardnesses were observed. The questions were still too wordy and respondents often interrupted before the questions were complete. Another small pilot of a moderately revised questionnaire was tested and this produced the following questions that became part of the standard questionnaire in 1982:

These next questions are about health care anyone in the family may have received during the 2 weeks outlined in red on that calendar.

1. During those 2 weeks , how many times did – see or talk to a medical doctor? Include all types of doctors, such dermatologists, psychiatrists, and ophthalmologists, as well as general practitioners and osteopaths. (Do not count times while an overnight patient is in the hospital).
2. (BESIDES THE TIME(S) YOU JUST TOLD ME ABOUT) During those 2 weeks did anyone in the family receive health care at home or go to a doctor's office, clinic, hospital or some other place? Include care from a nurse or anyone working with or for a medical doctor. (Do not count times while an overnight patient is in the hospital).
3. (BESIDES THE TIMES YOU HAVE ALREADY TOLD ME ABOUT) During those 2 weeks, did anyone in the family get any medical advice, prescriptions, or test results over the PHONE from a doctor, nurse, or anyone else working with or for a medical doctor?

Since 1982, these questions and others in the NHIS have come under intense scrutiny, largely as part of the work in the cognitive laboratory established by NCHS, and the questionnaire has undergone considerable revisions. Pilot surveys and formal comparative field tests of new questionnaires remains a vital part of the questionnaire process at NCHS, but they have been supplemented by focus groups and cognitive interviews in the laboratory using techniques such as protocol analysis and paraphrasing designed to improve comprehension, retrieval, judgment and communication.

For a description of problems with the 1982 NHIS Survey instrument and related questions on another health survey see Loftus, et al. (1992) and Suchman and Jordan (1992).

## V. Integrated Models For and Reviews of Nonsampling Error

A milestone in this effort to understand and model non-response errors was the development of an integrated model for sampling and non-sampling error in censuses and surveys, in connection with planning for and evaluation of the 1950 census (Hansen, Hurwitz, Marks, and Mauldin, 1951). This analysis-of-variance-like model, or variants of it takes the notion of Mean Squared Error = Variance + (Bias)<sup>2</sup> and breaks both the variance and the bias into components due to different sources of error.

Efforts to bring nonsampling error under control need to consider all of the components and such an approach is often referred to as *total survey design*. A single-minded approach focused on one source of error can bring it under control, but at the expense of ignoring other sources. For example, the traditional standardized survey interviewing approach described in Fowler (2002) and in Fowler and Mangione (1990) is designed to minimize interviewer related error, but it often does this at the expense of increased measurement errors associated with the respondents and not the interviewers.

The basic approach of using integrated models and total survey design has served as the basis of much of the work on non-sampling error over the past 35 years, especially in federal statistical agencies such as the Bureau of the Census, in a number of university based survey research centers and in a small handful of commercial survey organizations.

An excellent qualitative application of the error structure of the Current Population Survey is given in Brooks and Bailer (1978) and in an updated form in CPS Technical Paper 63 (2000), issued jointly by the Bureau of Labor Statistics and the Bureau of the Census. Reviews of the non-sampling error literature are given by Mosteller (1978) and Fienberg and Tanur (1983). Groves (1989) and Lessler and Kalsbeek (1992) give extended treatments of updated approaches and variants of this census model, making a careful distinction between random and fixed components from the various sources of error.

## VI. Examples

In the case studies we shall examine over the remainder of the semester we will spend considerable amounts of time focusing on issues of non-sampling error and the efforts of investigators to measure it and bring it under control.

## References

Brooks, C. A. and Bailar, B. A. (1978). *An Error Profile: Employment as Measured by the Current Population Survey*. Statistical Working Paper 3, Office of Federal Statistical Policy and Standards. Washington: U. S. Department of Commerce.

Bureau of Labor Statistics and U.S. Bureau of the Census (2002). Design and Methodology. Current Population Survey Technical Paper 63, Washington, DC. An updated version is available at [www.bls.census.gov/cps/tp/tp63.htm](http://www.bls.census.gov/cps/tp/tp63.htm).

Deming, W. E. (1944). On errors in surveys. *American Sociological Review*, **19**, 359-369.

Fienberg, S. E. and Tanur, J. M. (1983). Large scale social surveys: perspectives, problems, and prospects. *Behavioral Science*, **28**, 135-153.

Fowler, F. J., Jr. (2002). *Survey Research Methods*. (3rd Edition) Newbury Park, CA: Sage.

Fowler, F. J., Jr., and Mangione, T. W. (1990). *Standardized Survey Interviewing : Minimizing Interviewer-Related Error*. Newbury Park, CA: Sage.

Groves, R. M. (1989). *Survey Errors and Survey Costs*. New York: 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.

Lessler, J. T., and Kalsbeek, W. D. (1992). *Nonsampling Error in Surveys*. New York: Wiley.

Loftus, E.F., Smith, K.D., Klinger, M.R., and Fielder, J. (1992). Memory and mismemory for health events. In J.M. Tanur (ed.), *Questions About Questions*. New York: Russell Sage Foundation, 102-137.

Mosteller, F. (1978). Errors: I. Nonsampling errors. In *International Encyclopedia of Statistics*, W. H. Kruskal and J. M. Tanur, eds. New York: Macmillan and the Free Press, pp. 208-229.

Payne, S. L. (1951). *The Art of Asking Questions*. Princeton: Princeton University Press.

Plewes, T. J. (1994). Federal agencies introduce redesigned Current Population Survey. *Chance*, **7**, 35-41.

Schuman, H. and Presser, S. (1981). *Questions and Answers in Attitude Surveys*. New York: Academic Press.

Statistical Policy Working Paper 10. (1983). *Approaches to Developing Questionnaires*. Federal Committee on Statistical Methodology, Statistical Policy Office, Office of Management and Budget, Washington, DC.

Suchman, L. and Jordan, B. (1992). Validity and collaborative construction of meaning in face-to-face interviews. In J.M. Tanur (ed.), *Questions About Questions*. New York: Russell Sage Foundation, 241-267.

Tanur, J. M., ed. (1992). *Questions About Questions. Inquiries into the Cognitive Bases of Surveys*. New York: Russell Sage Foundation.