Journal of Official Statistics
Vol. 3, No. 4, 1987, pp. 323–325

Nonsampling Errors

Barbara A. Bailar[1]

1. Introduction

Nonsampling errors are ubiquitous. These errors are generated by reporting forms, unclear definitions or concepts, data collection procedures, interviewers, respondents, coders, analysts, estimation procedures, maps, seasonal adjustment, analysis, and interpretation of results. In fact, nonsampling errors include all errors except the error resulting from using a sample rather than the entire population. Nonsampling errors affect all statistical activities involved in producing statistics, including censuses, surveys, and administrative record operations. Unlike sampling error, statisticians rarely have a single number to attach to a statistic that measures the additional error.

Nonsampling errors include both biases and variances that arise in almost every phase of a statistical operation. The effect of these biases and variances is far greater than the effect of sampling error, especially for statistics at national levels where sampling error is usually held to levels that are appropriate for the uses of the data. Nonsampling errors are studied in unsystematic ways, with a few errors receiving much attention, many receiving no attention, and the interaction between these errors receiving very little consideration. Comprehensive studies are rarely carried out because the costs are high; changes in procedure may be necessary which in turn raises costs again, and sometimes because people believe that looking for and publicizing error is wrong. Nevertheless, it is my experience that when carefully planned studies are carried out, they usually contribute to improving statistical systems. Two studies of this type have had significant consequences. One study on recall by Neter and Waksberg (1965) is still cited as a major work on the length of recall and reference periods; the second is the interviewer variance study of 1950 that led to the use of self-enumeration in subsequent U.S. censuses. (See Hansen, Hurwitz, and Bershad (1961).)

There are few studies that have actually contributed to our understanding of nonsampling errors. Of the studies that have been conducted, many have produced contradicting results. Often the topics have been too narrow and the documentation has been poor, if attempted at all. Nor are these studies used to develop or test a theory of nonsampling error. There is more folklore in currency than supportable generalizations.

2. Kinds of Nonsampling Errors

Before starting a systematic study of nonsampling errors, statisticians should categorize these errors according to their sources. This categorization is important for a systematic reduction of these errors. It is important to make effective trade-offs of resources, based not only on the sources of the errors, but also on the likely impact of different kinds of errors

[1] U.S. Bureau of the Census, Washington, D.C., U.S.A.

on the uses of the data. Nonsampling errors that result from data processing procedures demand a different kind of treatment from those caused by interviewers or respondents. Broad categories of nonsampling error sources are: setting objectives for a statistical activity, translating concepts into definitions, questionnaire design and wording, sample design, data collection procedures and materials, training of employees and interviewers, respondents, coders, keyers, design and implementation of data processing procedures, estimation techniques, and analysis and interpretation. Of course, not all of these sources exist for every statistical study. In addition to these sources, there are some complex interactions, such as those that arise between interviewers and respondents, that could be based on race, ethnicity, age, sex, or general perceptions of social class.

3. What Do We Know

There are some kinds of nonsampling errors that we know a great deal about, at least in certain settings. Some of these are: coverage, interviewer effects, some types of questionnaire effects, some types of recall-of-information effects, respondent reluctance, use of dependent verification, and nonresponse. Although there is clearly a lot more to learn, we do know such things as:

1. the relative coverage of censuses and surveys (Bailar (1986));
2. the identification of populations that we get heavy nonresponse from or tend to miss altogether (Bailar (1986));
3. that simple, short questionnaires get higher item response (Cole and Altman (1977));
4. that people remember more events when the recall period is shorter (Bushery (1981));
5. that people tend to place events as occurring more recently in time than when they actually occurred (Bailar and Biemer (1984));
6. that dependent verification of keying, coding, or other similar operations yields serious underestimates of error (Minton (1969)).

There are other sources of error about which we know little, and this small body of knowledge generates conflicting hypotheses. For example, there is the effect of dependent interviewing in continuing surveys where respondents can see or are told their previous answers. One hypothesis is that dependent interviewing leads to underestimates of real change. The other hypothesis is that respondents are less likely to report spurious changes, and estimates of change are better and more consistent. At the Census Bureau, each of these hypotheses has its supporters. This results in dependent interviewing being used in some surveys while not in others. These conflicting views determine the procedures for many surveys, though both are, in fact, still untested hypotheses. Neither hypothesis is supported by measurable evidence. We cannot say for sure what impact dependent interviewing has on the survey data.

Though statisticians often know a great deal about types of people or companies that do not respond, they often know little about how nonresponse and nonresponse adjustments affect their data. Different statistical activities use different kinds of adjustment procedures, sometimes imputing missing data for high percentages of the population. In some cases industry averages, known to be wrong for smaller companies, are imputed. In addition, variance estimates are frequently made on the basis of imputed data, creating a downward bias in the estimates.

There are other areas where we need to develop measures to estimate the effects of certain treatments – for example, the training of survey personnel. At present, we have very little idea of the benefits of additional training, such as adding an extra day, using videos, or the importance of seeing live interviews. It seems likely that a variety of training techniques, which reinforce each other, are superior to

simply reading material aloud to a class. But how do we measure that effect? By the number of people who complete training? By the number who complete the job? By the number of nonresponses? By the number of incorrect answers? It would be useful to be able to show that one day's additional training or using a varied training package resulted in less interviewer turnover and higher quality data.

If we had more documented studies with measurable outcomes, we could go a long way in improving statistical output.

4. Where Do We Go From Here

For years, George Box and others have advocated the use of EVOP (evolutionary operations) as an ongoing mode for the gradual improvement of industrial processes. By using well-tested experimental designs in ongoing operations, not making large changes in process variables, and to limiting the size of samples where changes in categorical variables would be tested, the effects of small changes can be measured. This same process needs to be built into statistical operations such as surveys and censuses. By building experiments into every survey, census, or administrative data operation, we would soon establish a systematic body of knowledge. These studies would build on what was learned earlier, so that new information and especially information on interactions would be produced.

To provide the impetus for undertaking EVOP, statisticians must do more to illustrate the impact on nonsampling errors on uses of data. For example, we now have a better understanding of the impact of undercounting in a census on uses of census data; we also know that in the U.S., foreign trade export data is more seriously understated than imports, making for over-estimates of deficits in the balance of trade. When users understand these impacts on important uses, they usually support, sometimes demand, research on improved techniques. When we start incorporating controlled studies routinely into existing surveys and censuses, and then act on the result to improve procedures, we will be well on the way to moving survey- and census-taking from an art to a science.

5. References

Bailar, B.A. (1986): Information Needs, Surveys, and Measurement Errors. Presented at the International Symposium on Panel Surveys, November, 1986. To be published in symposium volume.

Bailar, B.A. and Biemer, P.P. (1984): Some Methods for Evaluating Nonsampling Error in Household Censuses and Surveys. In W.G. Cochran's Impact on Statistics, edited by P.S.R.S. Rao and J. Sedransk, John Wiley, New York, pp. 253–274.

Bushery, J.M. (1981): Recall Biases for Different Reference Periods in the National Crime Survey. American Statistical Association, Proceedings of the Section on Survey Methods Research, pp. 238–243.

Cole, S.J. and Altman, M.R. (1977): An Experiment with Short Forms in the Annual Survey of Manufactures. Presented at the meeting of the American Statistical Association. Proceedings of the Business and Economics Statistics Section, pp. 259–262.

Hansen, H.H., Hurwitz, W.N., and Bershad, M.A. (1961): Measurement Errors in Censuses and Surveys. Bulletin of the International Statistical Institute, 38, pp. 359–374.

Minton, G. (1969): Inspection and Correction Error in Data Processing. Journal of the American Statistical Association, 64, pp. 1256–1275.

Neter, J. and Waksberg, J. (1965): Response Errors in Collection of Expenditures Data by Household Interviewers: An Experimental Study. Bureau of the Census Technical Paper No. 11.

Received November 1987