

Non-sampling Errors: Some Reflections

*Thomas B. Jabine*¹

I was first drawn to survey work when I was exposed to what seemed to me a remarkable fact: that one could make inferences about a population, with a known degree of confidence, from a properly selected sample of its members. I did a master's thesis on sampling and then decided to take a job at the U.S. Census Bureau, where exciting developments were taking place in the theory and application of finite population sampling techniques. It was a fortunate decision. I joined a remarkable group of statisticians and gained valuable experience in the uses of sampling in surveys and censuses.

I also soon learned that although one could, with sufficient care and ingenuity, keep the level of sampling errors under control, there was a much more insidious influence that threatened the quality of survey data. It was called *non-sampling error*. Unlike the relatively well behaved and predictable sampling errors, non-sampling errors could strike at any stage of a survey and took many forms. Variable types of non-sampling errors could be measured by carefully designed (and costly) studies, but the dreaded *bias* was especially unpredictable, hard to detect, and difficult to eliminate. I reached the conclusion that one could not take responsibility for the quality of survey results without allocating a major share of one's attention to the measurement and control of non-sampling errors. Over the

years my interests have shifted more and more towards non-sampling errors.

From this perspective, I will present some views on non-sampling errors. How prevalent is the malady? Can a cure be found or at least some reasonably effective treatments? My comments are based primarily on U.S. censuses and surveys.

There is fairly good agreement among statisticians about the definition of non-sampling errors, although there are some fuzzy areas where sampling and non-sampling errors interact. The U.S. Census Bureau survey model introduced by Hansen, Hurwitz and Bershad (1959), partitions non-sampling error into several components. Within this framework, we can attempt to determine the relative importance of different types and sources of non-sampling errors and allocate error control measures where the benefit appears to be greatest. When sampling errors are included in this process, we speak of total survey design, a concept first introduced by Dalenius (1974).

Applicability of error models and ability to measure and control the components of non-sampling error are affected by the choice of survey variables. The measurement of bias for a survey variable depends on the ability to define and measure a true value for that variable. When we try to measure attitudes and perceptions in surveys, it is difficult, perhaps impossible, even to define true values, let alone measure them. Nevertheless, we have all seen attitude survey results which we believe are biased in some sense, whether by the

¹ Statistical Consultant, Washington, D.C., U.S.A.

sponsorship of the survey, the wording and context of the questions or some other feature of the survey.

Measurement of subjective phenomena – opinions, attitudes, perceptions, and knowledge – is a particularly difficult undertaking. Some responsible practitioners are doing careful work in the collection and presentation of “soft” data. But there are also innumerable efforts of poor quality, either because the surveyors lack professional skills or, even worse, because their goal is not to learn the truth but to produce results that will support a particular point of view.

Even for “hard” data, survey research is constantly pushed to and sometimes beyond its limits. Many important things are extremely difficult to measure accurately: the distribution of wealth, the environmental exposure of individuals to carbon monoxide or the prevalence of the AIDS virus. Surveys on these topics must be conducted, but we need to be conscious of our limitations and try to ensure that sponsors and the public are aware of these limitations. Statistical agencies in the United States have developed standards for presenting information about errors in survey data (U.S. Bureau of the Census (1974); Energy Information Administration (1985); Center for Education Statistics (1987)).

At the other end of the spectrum are market research surveys that provide data used for advertising. We frequently encounter advertising claims based on survey data. These claims are seldom accompanied by information that would permit even an informal judgment about the accuracy or reliability of the survey data. Somewhere in the middle are the better-known opinion surveys whose results appear in the media, often accompanied by a rough indication of the level of sampling error, a very brief account of the survey design and (highly desirable for opinion surveys) the exact wording of the questions used.

How much do sponsors, other users, and

the public care about non-sampling errors? This depends largely on where their interests lie. Sponsors who want to prove a particular point may insist on biased designs if the results of an objective study are not to their liking. Insistence on full disclosure of survey design and procedures is the only protection in such cases. If disclosure is not demanded, dishonest sponsors will continue to obtain and exploit the predetermined results.

For the sponsors and survey organizations that are trying to conduct reasonably objective surveys, the attention given to non-sampling errors will depend on their professional abilities, the resources available and the feedback from users and the public. When user interests are affected by the quality of the data, users will be heard from. One prime example in the United States relates to undercoverage in the population census (Norwood (1987)). User concerns about the effects of differential undercoverage on various population groups have led to a vigorous debate about whether the 1990 census counts should be adjusted, a controversy which has not yet been fully settled. On the other hand, there has been relatively little pressure for the government to revise the monthly estimates of unemployment from the current population survey, even though reinterviews have consistently shown that there is a response bias in the direction of under-reporting.

Another example of strong user feedback occurred in connection with the 1983 Survey of Consumer Finances, which sought to measure the distribution of wealth in U.S. households. The initial report of the survey showed that the share of wealth held by the top 0.5 percent of households had increased from 25 to 35 percent between 1963 to 1983. Much was made of these findings by those who argued that the present U.S. administration had been favoring the wealthy. It was subsequently discovered that a single respondent whose sampling weight happened to be large had

been incorrectly credited with owning a business worth \$200,000,000 when the correct value was \$2,000,000. When this error was corrected, the 1983 share of the top 0.5 percent of households came to 27 percent. This episode led to considerable embarrassment for those who sponsored and conducted the survey and for the users who rushed to judgment based on the initial results (Washington Post (1986)).

I have discussed this last example of non-sampling error, which might be characterized as *catastrophic*, with my colleagues and some of them have responded with similar experiences. For example in which the incorrect treatment of one or a few reporting units led to substantial errors which were not detected prior to publication. This experience leads to the following thoughts:

- High-impact observations should always be identified and carefully reviewed prior to the release of data. If they turn out to be correct and have large weights, the problem becomes difficult theoretically. One must consider various estimation procedures that have been proposed for dealing with outliers (Woodruff (1963) and Chambers (1986)).
- When such an error occurs, the first reaction is naturally to ask how to correct it with minimal embarrassment. The next reaction should be to ask how the error occurred and what can be done to prevent similar catastrophic errors in the future. The crusading spirit and process control that characterize today's quality control can help to do this, especially for periodic surveys.

Measurement and control of non-sampling errors demands and will continue to demand major attention and effort by survey researchers. The problems will not disappear and we will continue to seek information on new and more complex topics in surveys. Without presuming to prescribe a comprehensive program of treatment, I would like to suggest

some possibilities for preventing the non-sampling error malady from reaching epidemic proportions:

- Show more consideration for respondents. Try to avoid asking them to do things that are difficult, impossible, embarrassing, or of little interest to them. Include cognitive scientists on the survey research team (Jabine et al. (1984)). This may lead to a better understanding of respondents' points of view and how they react to the tasks they are asked to perform. Better treatment of respondents can bring both immediate and longer-term gains.
- Promote the application of modern quality control and improvement techniques and philosophy to survey programs, especially for periodic surveys, and emphasize process control and the participation of everyone involved in the design and conduct of surveys.
- Promote the education of survey personnel in the measurement and control of non-sampling errors and the education of potential data users on the nature and implications of non-sampling errors. Appropriate training materials are needed, both for formal training in universities and on-the-job training.

Finally, it seems to me essential that members of the profession take some responsibility not just for their own surveys but for the survey enterprise as a whole. It is not enough to say that you are doing the best you can to do good quality work in your own surveys. Surveys that are useless, of low quality or dishonest are a threat to the entire profession; they will diminish both the general level of cooperation by potential respondents and the public's trust in survey findings. We must all help to expose and discourage substandard practices.

References

- Chambers, R. (1986): Outlier Robust Finite Population Estimates. *Journal of the Amer-*

- ican Statistical Association, 81, pp. 1063–1069.
- Dalenius, T. (1974): *Ends and Means of Total Survey Design*. Stockholm: University of Stockholm.
- Hansen, M., Hurwitz, W., and Bershad, M. (1959): *Measurement Errors in Censuses and Surveys*. *Bulletin of the International Statistical Institute*, 38, pp. 359–374.
- Jabine, T., Straf, M., Tanur, J., and Tourangeau, R. (Eds.) (1984): *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines*. National Academy Press, Washington, D.C.
- Norwood, J. (1987): *What is Quality?* Proceedings, Third Annual Research Conference, Bureau of the Census, Washington, D.C.
- U.S. Bureau of the Census (1974): *Standards for the Discussion and Presentation of Errors in Data*. Technical Paper 32, Department of Commerce, Washington, D.C.
- U.S. Center for Education Statistics (1987): *Standards and Policies*. Department of Education, Washington, D.C.
- U.S. Energy Information Administration (1985): *Energy Information Administration Standards Manual*. Department of Energy, Washington, D.C.
- Washington Post (1986): *Report Overstated Gains of "Super Rich"*. August 22.
- Woodruff, R. (1963): *The Use of Rotating Samples in the Census Bureau's Monthly Surveys*. *Journal of the American Statistical Association*, 58, pp. 454–467.

Received December 1987