

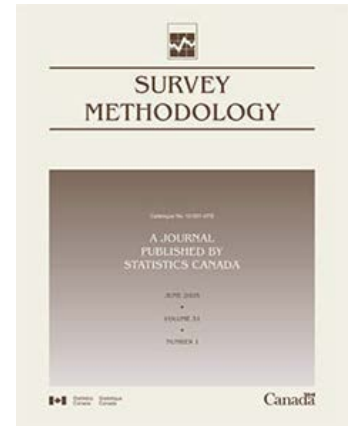
Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

Comments on the Rao and Fuller (2017) paper

by Sharon L. Lohr

Release date: December 21, 2017



Statistics
Canada

Statistique
Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

email at STATCAN.infostats-infostats.STATCAN@canada.ca

telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following toll-free numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-877-287-4369

Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under “Contact us” > “Standards of service to the public.”

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Standard table symbols

The following symbols are used in Statistics Canada publications:

- . not available for any reference period
- .. not available for a specific reference period
- ... not applicable
- 0 true zero or a value rounded to zero
- 0^s value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
- ^P preliminary
- ^r revised
- X suppressed to meet the confidentiality requirements of the *Statistics Act*
- ^E use with caution
- F too unreliable to be published
- * significantly different from reference category ($p < 0.05$)

Published by authority of the Minister responsible for Statistics Canada

© Minister of Industry, 2017

All rights reserved. Use of this publication is governed by the Statistics Canada [Open Licence Agreement](#).

An HTML version is also available.

Cette publication est aussi disponible en français.

Comments on the Rao and Fuller (2017) paper

Sharon L. Lohr¹

Abstract

This note by Sharon L. Lohr presents a discussion of the paper “Sample survey theory and methods: Past, present, and future directions” where J.N.K. Rao and Wayne A. Fuller share their views regarding the developments in sample survey theory and methods covering the past 100 years.

Key Words: Data collection; History of survey sampling; Probability sampling; Survey inference.

Rao and Fuller deserve thanks for their succinct review of a field to which they both have contributed so much. It is no small feat to summarize the history of probability sampling and outline future directions in 16 pages!

It is always hazardous to predict the future. But reviewing the history of survey sampling allows us to see how the pioneers of the field dealt with the challenges of their day, and how those challenges and their solutions relate to today’s issues.

To begin, let us compare the advantages and disadvantages of probability sampling today with the advantages and disadvantages that were perceived in the middle of the twentieth century, when probability samples were starting to have widespread use. The following lists are derived from Parten (1950, Chapter 4); the early sampling books by Deming (1950) and Hansen, Hurwitz and Madow (1953a) have similar descriptions. Parten’s advantages of probability samples, relative to taking a census of the population or taking a convenience sample, are of four types:

- A1. Estimates can be obtained faster from a sample than from a census. Fewer interviews are needed and, in the 1950s, data processing and tabulation could be done faster for a small data set than for a large one. Parten wrote: “This time-saving advantage is especially important in studies of our modern dynamic society. Conditions change so rapidly that unless short-cut methods are devised for measuring social situations, the measurement is out of date before the survey or poll is completed.” (Parten, 1950, page 109).
- A2. Estimates from a sample are less expensive than a census because fewer interviews are needed. This translates into lower costs for field staff and training.
- A3. The survey can be tailored to the estimates of interest. The sampler can be more careful in the data collection, asking exactly the questions wanted and taking steps to minimize bias from nonresponse and other sources. A census, by contrast, may have few questions and limited opportunity for follow-up.
- A4. Probability sampling allows the sampler to design the survey to achieve a desired precision and later report the achieved precision, without relying on model assumptions. Deming (1950, page 10) emphasized that not only can the sampling errors be calculated from probability

1. Sharon L. Lohr, Arizona State University, Tempe, AZ. E-mail: sharon.lohr@asu.edu.

samples, but that “the biases of selection, nonresponse, and estimation are virtually eliminated or contained within known limits.” Hansen et al. (1953a, page 10) stated: “With probability sampling methods one can get away completely from dependence upon judgment for determining precision. Under these circumstances, and with reasonably large samples, the precision of the results from the sample can be measured from the sample itself.”

Parten also reviewed the disadvantages of taking a sample rather than a census:

- D1. It is difficult to do sampling well, and to obtain representative samples. Mistakes in following the sampling protocol may introduce errors into the estimates, and results can be misleading if a sample is designed or analyzed badly. Additionally, a shortage of experienced survey statisticians makes it difficult for the would-be survey-taker to obtain technical assistance.
- D2. The small size of the sample limits the information that can be obtained. Rare subpopulations have few observations in a sample. Additionally, the number of cross-tabulations is limited because there are too few cases in some subclassifications of interest.

How do Parten’s advantages and disadvantages of probability samples hold up today? The disadvantages still exist. In particular, the demand for more detailed, more up-to-date, and more comprehensive information is increasing every year (D2). But while surveys still have the advantage (A3) that they can be tailored to answer the questions of interest, advantages (A1) and (A2) have diminished. In the 1950s, it was often expensive to collect *any* type of data. Even data from a small convenience sample could require expensive-to-collect interviews or labor-intensive transcription of paper records. But today, huge convenience samples can often be obtained with much less cost, while probability samples such as the American Community Survey or the National Crime Victimization Survey are increasingly expensive as response rates continue to decline. The convenience sample information may also be available faster than data from a high-quality probability sample, which requires months to weight the data, compute estimates, and perform quality checks.

The advantage (A4) of being able to plan for a desired precision has also diminished. Most large surveys still use design-based methods to report the precision of the survey. But the design-based margin of error that is reported generally includes only the sampling error and has the implicit assumption that the survey weighting has removed the nonresponse bias from the estimates. As response rates decrease, there is increasing reliance on judgment, through the use of model assumptions, to determine precision.

The landscape for data collection is thus different than it was in the 1930s, 1940s, and 1950s when probability sampling techniques were developed and implemented. Then, probability sampling in the United States answered an urgent need for faster and cheaper information about agriculture, business activity, manufacturing output, characteristics of the labor force, and other social and economic indicators. The pioneers of survey sampling methods revolutionized data collection during this period. Duncan and Shelton (1978) argued that this revolution was made possible by parallel developments in statistical theory, national income and product accounts, computing capacity, and organization of the statistical system.

Although the available data sources, infrastructure, technology, and methods have changed, the main problem facing us today is the same as in 1950: How can we best collect and make inferences from data to inform policy and research questions? If the current framework of probability samples did not exist, and we were tasked with constructing a system of collecting data, what would we do? For many problems, we would want to build a data collection system that is modular and can adapt to new sources of data and new technology for data collection. Much of the methodology that Rao and Fuller reviewed would be useful for this system, but new infrastructure and methodology—and perhaps another revolution—are needed.

As an example, consider the U.S. National Automotive Sampling System (NHTSA, 2017a). The system has two component surveys. The first component is a stratified multistage probability sample of 50,000 to 60,000 police accident reports (PARs) from the universe of approximately 6 million annual PARs, where PARs from serious crashes are sampled with higher probabilities than PARs from crashes involving only minor property damage. Data elements from the sampled PARs are coded into the electronic database; no information external to the PAR is obtained. The second survey is a smaller probability sample of about 5,000 PARs with much more labor-intensive data collection, where specially trained crash investigators visit the crash scene, inspect the vehicle(s) involved in the crash, get permission to access medical records, interview witnesses, and obtain other detailed information about the crash. The data from these two surveys are used to investigate time trends in vehicle crashes and effects of vehicle features on traffic safety (see, for example, NHTSA, 2017b), and are used in thousands of research papers.

But suppose we were asked to design this data collection system afresh. I want to stress that these suggestions are my fantasy, and have no connection to any plans for the surveys, which are constrained by current practical considerations and budgetary limitations to have a multistage sampling structure. If the current system did not exist, would it not be desirable to design the first survey to take a census of PARs instead of a sample? This task would not necessarily be easy. Hetzel (1997) described the long and laborious process by which the United States established the Vital Statistics System, requiring cooperation from state and local government agencies, uniform data collection procedures, and intensive research to validate the accuracy and coverage of records. Obtaining a census of PARs would similarly take huge initial investments to develop infrastructure and to secure cooperation from states and police jurisdictions. After that investment, however, the data collection would be established and PARs could be transmitted electronically as they were collected or updated.

The advantages of having a census of PARs instead of a sample would be numerous. One advantage would be that statistics would be available much more quickly, since one would not need to wait until the end of the data collection year to weight and publish the survey data: statistics could be updated as the data came in. The largest advantage of a census, however, would be the extra information on subpopulations. This would allow better monitoring of the data to detect potential safety hazards. In a sample of size 50,000, a particular make/model/year of vehicle may be represented by only a handful of observations (if any); the sample size would be much larger in the census. In some surveys, as Rao and Fuller point out, small area estimation methods can be used to model results for subpopulations with small sample sizes. But for crash data, often a subpopulation is of interest because it is suspected to be an outlier—it is suspected there are more crashes for a certain car make or vehicle feature than would be predicted by a model. These outliers

cannot be detected from a small area model. The only way to obtain information on potentially outlying subpopulations is to collect more data on them.

With a census of PARs, though, where do survey research methods come in? There would inevitably be missing data that would need to be investigated and modeled, and a two-phase design might be used to obtain information from nonresponding states or police jurisdictions. But the main survey design problem would be for two aspects: first, sampling could be used to audit the census of accident reports, and second, sampling would be needed for the labor-intensive crash investigation part of the system. The census of PARs would provide a rich sampling frame for the crash investigation system and other investigations. That rich frame information could be exploited in the sample design, possibly by using balanced sampling or a sampling design that can be dynamically adapted to data needs and to the continuously updated frame.

Of course, even a census of PARs might be outdated or insufficient for data needs in the future. As more vehicles are equipped with cameras and sensors, or as self-driving vehicles and surveillance systems become more prevalent, a sample or census of PARs may be replaced or supplemented with passively collected data. Increased use of large-scale passive data sources raises serious issues about privacy and data ownership, requiring much debate and research, and these issues are beyond the scope of this discussion. But, beyond the societal questions about the ethics of data collection, what new statistical methodology is needed to deal with the revolution in data availability?

I see three major areas of interconnected research needed for the short-term future, and these are related to the research problems that faced Parten, Deming, and Hansen in the middle of the last century.

- Better measures of uncertainty for survey estimates. When Hansen and Hurwitz (1949, page 365) wrote about the superiority of probability samples over judgment samples, they emphasized that the assumption-free nature of inference from probability samples depends on achieving high response rates: “In the Census Bureau it is usually assumed that if the required information is obtained from more than 95 per cent of the designated households one is entitled to feel fairly secure in assuming that the sample was taken in conformance with sampling theory, even though assumptions may be necessary for the remaining 5 per cent. It has been found that for some purposes trouble arises even when making assumptions for only 5 per cent.” Deming (1950, page 13) also used 95 percent as the lower bound for the validity of inference from probability samples: “A sample that is 95 or 98 percent a probability-sample and the other 5 or 2 percent a judgment-selection or judgment-adjustment for refusals, for people not at home, etc., may still be an excellent sample, although it is important to investigate the remaining 5 or even 2 percent as soon as possible.”

Over the years, the 95 percent threshold for using probability sampling methods for inference has drifted downward, to the point that now the same weighting methods are used for a sample with a response rate of 10 percent as for one with a response rate of 95 percent. As response rates have decreased, increasingly strong model assumptions have been made about response mechanisms and the undercovered and nonresponding populations, but uncertainty about these assumptions is generally not reflected in the reported confidence intervals: these are still primarily based on the sampling error. Lohr and Brick (2017) ascribed recent polling failures to the systems used to

derive confidence or posterior prediction intervals, and argued that new statistical methods are needed for reporting interval estimates that better reflect uncertainty about the estimates. As Parten (1950, page 403) said: “It is not unusual to find very refined statistical techniques used to measure random errors in data which are so biased that all the corrective devices known would not enable the surveyor to determine what the correct results should be.”

- Combining multiple sources of data. Methods such as record linkage, multiple frame surveys, and hierarchical models can facilitate combining data, and can also be used to evaluate data quality from different sources. Lohr and Raghunathan (2017) reviewed statistical methods for combining data from different sources, and argued that using multiple sources of information could also help with the problem of evaluating and incorporating bias errors into uncertainty estimates, although the statistical methods reviewed do not solve the problem of how to obtain estimates for subpopulations that may be missing from all sources.

While probability sampling methods have served society well for the last 70 years, they may play a more limited role in the future, perhaps being used in some data collections to validate and check other data sources instead of serving as the primary data sources themselves. Hansen, Hurwitz and Pritzker (1953b) viewed the census post-enumeration surveys, in which intensive efforts were made to determine the most accurate information possible from an area sample, in this light. They also wrote about the need to weigh the costs of obtaining accurate statistics against the option of obtaining larger sample sizes or more variables. Hansen, Madow and Tepping (1983) argued that it is precisely at times of societal change that the unbiasedness guaranteed by probability sampling is essential, and targeted use of high-quality probability samples can provide assessments of the coverage and accuracy of information from other data sources.

- Finally, there is a need for a renewed focus on design, which dominated much of the literature between 1920 and 1960. A modular data collection system, relying on different data sources for different information needs, could adapt to changing societal needs and new technology for data collection. Designs are needed that are robust to errors in individual sources and allow assessment of those errors, and are also robust to potential changes in the data sources.

As Rao and Fuller pointed out, the statistical research community has repeatedly met the information needs of society through new innovations. The challenges of dealing with new data sources and missing data are great, but so were the problems faced in the past that led to the development of probability sampling, small area estimation, replication variance estimation, and imputation theory. The next revolution in sampling may be just around the corner.

Acknowledgements

Part of this discussion was adapted from the author’s 2016 JPSM Distinguished Lecture “The Essential Survey Statistician,” available at <https://www.jpsmclasses.umd.edu/Mediasite/Catalog/catalogs/default>.

References

- Deming, W.E. (1950). *Some Theory of Sampling*. New York: Dover.
- Duncan, J.W., and Shelton, W.C. (1978). *Revolution in United States Government Statistics 1926-1976*. Washington, D.C.: U.S. Department of Commerce.
- Hansen, M.H., and Hurwitz, W.N. (1949). Dependable samples for market surveys. *Journal of Marketing*, 14, 363-372.
- Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953a). *Sample Survey Methods and Theory. Volume I: Methods and Applications*. New York: John Wiley & Sons, Inc.
- Hansen, M.H., Hurwitz, W.N. and Pritzker, L. (1953b). The accuracy of census results. *American Sociological Review*, 18, 416-423.
- Hansen, M.H., Madow, W.G. and Tepping, B.J. (1983). An evaluation of model-dependent and probability-sampling inferences in sample surveys. *Journal of the American Statistical Association*, 384, 776-793.
- Hetzel, A.M. (1997). *U.S. Vital Statistics System: Major Activities and Developments, 1950-95*. Hyattsville, MD: National Center for Health Statistics. Available from <https://www.cdc.gov/nchs/data/misc/usvss.pdf>, last visited May 5, 2017.
- Lohr, S.L., and Brick, J.M. (2017). Roosevelt predicted to win: Revisiting the *Literary Digest* poll of 1936. *Statistics, Politics, and Policy*, 8, 65-84.
- Lohr, S.L., and Raghunathan, T.E. (2017). Combining survey data with other data sources. *Statistical Science*, 32, 293-312.
- National Highway Transportation Safety Administration (NHTSA, 2017a). *National Automotive Sampling System (NASS)*. Available from <https://www.nhtsa.gov/research-data/national-automotive-sampling-system-nass>, last visited May 5, 2017.
- National Highway Transportation Safety Administration (NHTSA, 2017b). *Traffic Safety Facts, 2015*. Available from <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812384>, last visited May 17, 2017.
- Parten, M. (1950). *Surveys, Polls, and Samples*. New York: Harper & Brothers.