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

Survey Methodology

Authors' response to comments on "Trends and directions in sample survey theory and methods"

by J.N.K. Rao and Sharon L. Lohr

Release date: June 30, 2025

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 infostats@statcan.gc.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-514-283-9350

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, the Agency has developed standards of service which its employees observe in serving its clients. 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.

Survey Methodology, June 2025
Vol. 51, No. 1, pp. 141-143
Statistics Canada, Catalogue No. 12-001-X

Authors' response to comments on "Trends and directions in sample survey theory and methods"

J.N.K. Rao and Sharon L. Lohr[1]

Abstract

The discussants highlight promising research topics for improving the quality and granularity of estimates from surveys. We agree that continued research is needed to evaluate models used for inference, and suggest development of measures of model dependence.

Key Words: Data integration; Machine learning; Model-based inference; Small area estimation; Survey quality.

We would like to thank the discussants for their thoughtful and insightful comments.

All of the discussants agree that the future of survey sampling involves increased dependence on models – to obtain more efficient designs and estimators, and to attempt to adjust for nonresponse and other non-sampling errors. Opsomer, Han, and Uppala's "generalized design-based statistician" models the complete selection process to account for deviations from the design-based selection probabilities caused by non-sampling errors. Haziza focuses on machine learning procedures for estimating finite population parameters and outlines research needed when these predictions are used for response propensity estimation, imputation, or small area estimation.

Ranalli emphasizes the potential of models for both design and estimation. Balanced sampling designs attempt to improve efficiency and guarantee subgroup representation through "pre-calibrating" the survey design with respect to a vector of known population totals $\mathbf{X}$, and Ranalli highlights extending the number of balancing variables through incorporating linear mixed models (Breidt and Chauvet, 2012) and spatial information. We appreciate her perspective that calibration to a more general function m is not a competitor to generalized regression but an enhancement of it. Her recent work on data editing via calibration (Cantarella, Neri and Ranalli, 2024) provides a promising mechanism for taking advantage of calibration variables and functions that relate to some study variables but not others.

All three discussants also share our concern about the quality of estimates when model assumptions are not met. As Opsomer, Han, and Uppala point out, inference from a full-response probability sample does not rely on the validity of model assumptions, even when calibration models are employed to improve efficiency of estimates. When models are used to compensate for nonresponse or other nonsampling errors, however, the coverage achieved by confidence intervals does rely on model assumptions. Moreover, these assumptions cannot be fully tested because the true response propensities are unknown and study variables are unavailable for nonrespondents (Rao, 2011).

1. J.N.K. Rao is Distinguished Research Professor, Carleton University, Ottawa, Ontario, Canada, K1S 5B6. E-mail: jrao@math.carleton.ca; Sharon Lohr, Professor Emerita, Arizona State University. E-mail: sharon.lohr@asu.edu.

The inability to validate model-based inferences motivated the development of probability sampling, as design-based inferences in a full probability sample rely on minimal assumptions that are met in almost all populations of interest. From the earliest years, however, there has been concern about survey inferences that depend on modeling assumptions. In her 1956 presidential address for the American Statistical Association, Cox (1957) explored the statistical frontiers of the time, with emphasis on the "design of experiments and investigations [sampling] continent." She described some of the research problems affecting the common sampling design countries (such as the stratified random sampling country and the multi-stage sampling country) and wrote: "In all the sampling territory, there are many internal political and economic frontiers to be cleared. These sampling countries now have fair control over sampling errors but relatively little over non-sampling errors" (page 7). She was particularly concerned about inference in the presence of nonresponse (even though response rates for surveys such as the U.S. Current Population Survey then exceeded 95 percent) and asked: "How far are we justified in using statistical methods based on probability theory for the analysis of nonexperimental data?" (page 10).

Cox's question is still valid. How far are we justified in using statistical methods based on the survey design when models are used to compensate for the missing data? Weighting and calibration models for nonresponse are generally thought to reduce bias, but some empirical evaluations have found that the models decreased accuracy (Fladmoe and Bergh, 2022). In a similar vein, Buttice and Highton (2013) found large differences between direct subpopulation estimates obtained from large surveys and corresponding subpopulation estimates obtained from small surveys through multilevel regression with poststratification (MRP), concluding that "the conditions necessary for MRP to perform well will not always be met" (page 449). Continued theoretical and empirical research is needed to evaluate models relied upon for inference – particularly for models that have not been as well studied such as those involving machine learning.

Ranalli states that it is essential to conduct robust diagnostics and evaluate the quality of auxiliary variables when models are employed. Lohr (2022) suggested diagnostics and evaluations that can be conducted for data integration methods. Quality of auxiliary variables can be audited for a probability sample from the auxiliary data source. When models are used to produce estimates for subpopulations, as in small area estimation and MRP, it is important to ensure the quality of data and estimates across the subpopulations.

We wholeheartedly agree with the discussants that transparency is crucial, and appreciate the suggestion from Opsomer, Han, and Uppala to "more explicitly recognize this model dependence in the information provided with survey datasets". One possibility is to report measures of model dependence for estimates. Imputed datasets, for example, are often accompanied by a table listing the percentage of imputed observations for each variable. Some providers of small area estimates have published values of $\hat{\gamma}$ along with the Fay-Herriot (1979) estimates $\hat{\theta}_i = \hat{\gamma}_i \overline{y}_i + (1-\hat{\gamma}_i)\, \mathbf{x}_i^T \hat{\boldsymbol{\beta}},$ where $\overline{y}_i$ is the direct estimator from the survey and $\mathbf{x}_i^T \hat{\boldsymbol{\beta}}$ is the prediction from the regression model – here, $(1-\hat{\gamma}_i)$ measures the dependence on

the model prediction. Analogous measures of model dependence, and additional diagnostics, could be developed for nonresponse weighting models and MRP.

The discussants suggest numerous research ideas for improving the quality of statistics produced from survey samples. We look forward to seeing these ideas developed.

Acknowledgements

J.N.K. Rao's research was supported by a research grant from the Natural Sciences and Engineering Research Council of Canada.

References

Breidt, F.J., and Chauvet, G. (2012). Penalized balanced sampling. *Biometrika*, 99(4), 945-958.

Buttice, M.K., and Highton, B. (2013). How does multilevel regression and poststratification perform with conventional national surveys? *Political Analysis*, 21, 449-467.

Cantarella, M., Neri, A. and Ranalli, M.G. (2024). Estimating the distribution of household wealth: Methods for adjusting survey data estimates using national accounts and rich list data. *Review of Income and Wealth*, 70(3), 551-580.

Cox, G.M. (1957). Statistical frontiers. *Journal of the American Statistical Association*, 52(277), 1-12.

Fay, R.E., and Herriot, R.A. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74(366a), 269-277.

Fladmoe, A., and Bergh, J. (2022). The use of adjustment weights in voter surveys. Correcting for panel attrition and nonresponse can produce less accurate estimates of voting behavior. *Electoral Studies*, 78, 102486.

Lohr, S.L. (2022). Comments on "Statistical inference with non-probability survey samples". *Survey Methodology*, 48(2), 331-338. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2022002/article/00005-eng.pdf.

Rao, J.N.K. (2011). Impact of frequentist and Bayesian methods on survey sampling practice: A selective appraisal. *Statistical Science*, 26(2), 240-256.