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

Survey Methodology

# Preface to the special issue for papers presented at the 29th Morris Hansen Lecture on the use of nonprobability samples


by Partha Lahiri

Release date: June 25, 2024

## 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, 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.

Published by authority of the Minister responsible for Statistics Canada



**An HTML version is also available.**

*Cette publication est aussi disponible en français*.

Survey Methodology, June 2024
Vol. 50, No. 1, pp. 1-2
Statistics Canada, Catalogue No. 12-001-X


# Preface to the special issue for papers presented at the 29th Morris Hansen Lecture on the use of nonprobability samples

**Partha Lahiri**[1]

Neyman's seminal paper transformed survey sampling, leading to widespread adoption of probability sampling and associated design-based methods, particularly within national statistical offices. However, perfect implementation of design-based methods relies on a perfect sampling frame of the target finite population, well-designed samples with known non-zero selection probabilities, no nonresponse, no measurement errors, and the use of sampling weights to correct for unequal probabilities. Under these conditions, consistency of traditional design-based estimators and their variance estimators can be assured for large samples, irrespective of the validity of any model that may have been used to construct the estimators. For large sample sizes, the probability sampling approach is indeed attractive to survey practitioners because the same estimation procedure can be used to handle different kinds of outcome variables without the need to model them separately.

Probability sample surveys encounter challenges, including noncoverage, measurement errors, declining participation rates, and high costs. Conversely, nonprobability surveys like voluntary surveys gain traction due to their convenience and cost-effectiveness. In nonprobability sampling, the selection probability mechanism is unknown. Moreover, often selection probabilities are zero for a subset of finite population units. Thus, traditional design-based methods cannot be used to construct estimates or their uncertainty measures, and one needs to rely on models whose assumptions may not always be verifiable. There is now a growing interest in integrating non-probability data with probability surveys, aiming to mitigate these challenges and leverage the strengths of both approaches.

Due to the increasing importance of nonprobability surveys, the Morris Hansen Lecture committee decided to organize the 29th Morris Hansen Lecture on the topic of "Working with Nonprobability Samples: Assessing and Remediating Bias." The Washington Statistical Society inaugurated the Morris Hansen Lecture Series in 1990, supported by a grant from Westat. Subsequently, the National Agricultural Statistics Service (NASS) joined as a co-sponsor of the event, and since then, has consistently hosted the lecture series nearly every year in Washington, D.C.

Given the ongoing Covid pandemic, the 29th Morris Hansen Lecture was conducted as a virtual event on March 1, 2022. The committee extended invitations to Jean-François Beaumont, Courtney Kennedy, and Yan Li, three esteemed experts in the field of nonprobability surveys, to deliver lectures based on their recent research in this area. This special issue features revised versions of three papers presented in the 29th Morris Hansen Lecture event along with discussions and rejoinders.

The first paper authored by Kennedy, Mercer and Lau investigates the measurement issues associated with nonprobability opt-in surveys, frequently utilized to generate estimates for rare domains due to cost

1. Partha Lahiri, Joint Program in Survey Methodology, University of Maryland, 1218 Lefrak Hall, College Park, Maryland 20742, U.S.A. E-mail: plahiri@umd.edu.

considerations. Through an extensive benchmarking study, the authors identify population subgroups characterized by significant bias in opt-in surveys, attributing a portion of this bias to bogus responses. Their findings underscore the importance of scrutinizing errors arising from bogus responses in nonprobability surveys, emphasizing the need to address not only selection bias but also the issue of erroneous responses.

The second paper authored by Li examines the conditional exchangeability assumption, which serves as a pivotal assumption in propensity score-based adjustment methods. Specifically, Li explores the validity of the exchangeability assumption under various balancing scores and devises an adaptive balancing score aimed at achieving unbiased estimation of finite population means.

The third paper authored by Beaumont, Bosa, Brennan, Charlebois and Chu represents a significant advancement in the field of inverse probability weighting methods for nonprobability samples aimed at mitigating selection bias. Their research encompasses data integration techniques incorporating both parametric and Classification and Regression Trees (CART) methods, with particular emphasis on accounting for the probability sample design. Of note is their substantial focus on variable selection within the context of their proposed methodologies.

I extend my sincere gratitude to Jean-François Beaumont, Editor of *Survey Methodology*, for graciously agreeing to dedicate a special issue of *Survey Methodology* and inviting me to serve as its Guest Editor. I also express my appreciation to the Morris Hansen lecturers – Jean-François Beaumont, Courtney Kennedy, and Yan Li – for accepting my invitation to contribute papers to this special issue based on their presentations.

Furthermore, I am grateful toauthors of all three papers – Jean-François Beaumont, Keven Bosa, Andrew Brennan, Joanne Charlebois, Kenneth Chu, Yan Li, Courtney Kennedy, Andrew Mercer, Arnold Lau – as well asthe discussants of the papers – Vladislav Beresovsky, J. Michael Brick, Michael R. Elliott, Julie Gershunskaya, Jae Kwang Kim, Yonghyun Kwon, Takumi Saegusa, Aditi Sen, and Changbao Wu – for their insightful commentary on the papers. Their contributions have stimulated valuable discussion and enhanced the depth of the research presented.

I believe that the papers, discussions, and rejoinders will serve as an invaluable reference for future research in this dynamic and challenging field.

Partha Lahiri
Guest Editor