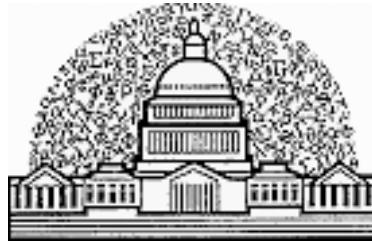


29th Annual Morris Hansen Lecture



Washington Statistical Society

Working with Non-Probability Samples: Assessing and Remediating Bias

Speakers: Courtney Kennedy

Director of Survey Research
Pew Research Center

Yan Li

Joint Program in Survey Methodology
University of Maryland

Jean - Francois Beaumont

Senior Statistical Advisor
Statistics Canada

Tuesday, March 1, 2022

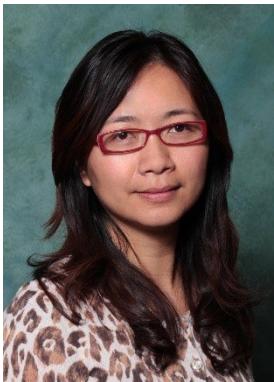
3:30 – 5:30 pm

The conference is being held virtually via Zoom



Exploring the Assumption That Online Opt-in Respondents Are Answering in Good Faith

Courtney Kennedy is director of survey research at Pew Research Center. In this role, she serves as the chief survey methodologist for the center, providing guidance to all of Pew Research Center's research areas. Her work has been published in *Public Opinion Quarterly*, *Journal of Survey Statistics and Methodology*, and *Journal of Official Statistics*. Her current research interests include data quality in online survey panels and approaches for weighting public opinion polls.



Exchangeability Assumption in Propensity-Score Based Adjustment Methods for Population Mean Estimation Using Non-Probability Samples

Yan Li is a Professor in the Joint Program in Survey Methodology (JPSM) and in the Department of Epidemiology and Biostatistics at the University of Maryland, College Park. She has served on various advisory committees, including the Design and Analysis Committee for the National Assessment of Educational Progress, ETS and NCES, and ASA Statistical Partnerships Committee Among Academe, Industry, and Government. She is currently the Methodology Program Chair for WSS. Prof. Li has served as a consultant to various federal agencies such as NCHS and NCI. Her current research interests focus on statistical method development for designing and analyzing complex samples, and for health disparity analysis. She has been an Associate Editor of *Sankhya B*. She is a Fellow of the American Statistical Association.



Reducing the bias of non-probability sample estimators through inverse probability weighting with an application to Statistics Canada's crowdsourcing data

Jean-François Beaumont is currently holding the position of Senior Statistical Advisor at Statistics Canada, where he has been working for more than 25 years. He is also the Editor of *Survey Methodology* and the president of the Survey Methods Section of the Statistical Society of Canada. Jean-François delivered many invited presentations and courses during his career, including the opening address of the Colloque francophone sur les sondages in 2018. His current research projects focus on statistical data

integration, small area estimation and other estimation issues with sample surveys.

Abstracts

Exploring the Assumption That Online Opt-in Respondents Are Answering in

Good Faith

Courtney Kennedy

Online opt-in (or nonprobability) samples have become widely, though not universally, adopted by U.S. public opinion researchers. Today, more than 80 percent of the surveys tracking key indicators of U.S. public opinion are conducted using online opt-in samples. Survey researchers recognize that in their raw, unadjusted form, responding samples often contain substantial biases that need to be reduced using statistical adjustment. Myriad approaches have been proposed, including quasi-randomization, superpopulation modeling, and doubly robust estimation. Most if not all of these adjustments rely on a key assumption: *that respondents are who they say they are*. But how often is that the case with opt-in samples? And are some types of respondents more prone to being insincere than others? Statistical adjustments can remedy the proportion of a subgroup in the sample being too low or too high. They can do this in sophisticated ways using numerous covariates and multivariate distributions. But statistical adjustments are less equipped to remedy a survey when, for example, substantial shares of respondents purporting to have a certain characteristic (e.g., a particular age, race or ethnicity) do not actually have that characteristic or answer in the affirmative regardless of what was asked. While companies selling opt-in samples have procedures to try to ensure that respondents are giving genuine answers, a number of studies show that bogus or insincere respondents still end up in survey samples. This presentation highlights findings from several Pew Research Center studies examining the sincerity and accuracy of online opt-in data. We focus in particular on subgroups where insincere responses are concentrated and errors (measured against federal benchmarks) are particularly large. We consider the magnitude of these errors relative to those observed in online probability-based samples as well as the implications for practitioners.

Exchangeability Assumption in Propensity-Score Based Adjustment Methods for Population Mean Estimation Using Non-Probability Samples

Yan Li

Along with the rapid emergence of web surveys to address time-sensitive priority topics, various propensity score (PS)-based adjustment methods have been developed to improve population representativeness for nonprobability samples. Exchangeability conditional on estimated PS is often assumed in various PS-based adjustment methods with PS estimated from propensity models that are fitted by the combined probability and nonprobability sample. Current literature makes mixed suggestions on estimating PS with or without probability sample weights to achieve the conditional exchangeability. In this talk, I will discuss different exchangeability assumptions under PS-based weighting and matching methods, and propose an adapted exchangeability assumption. Under this assumption, two innovative PS-based matching methods are proposed: 1) two-step matching, and 2) calibration for unbiased estimation of the population mean. Monte Carlo simulation studies are designed under varying degrees of departure from the adapted exchangeability assumption to evaluate the bias of the proposed estimates. Using data from National Institutes of Health and the American Association of Retired Persons (NIH-AARP Diet and Health Study, a nonprobability cohort sample)

and the US National Health Interview Survey (NHIS, a probability survey sample), the proposed methods will be applied to estimate the prevalence of self-reported diseases and all-cause or all-cancer mortality rates for people aged 50-71 in the US. To test our methods, we will purposely select outcome variables that are available in both the NIH-AARP and the NHIS. Thus, the amount of bias in NIH-AARP estimates corrected by the proposed methods can be quantified in practice, assuming the weighted NHIS estimates are (approximately) unbiased for the targeted 50-71 year olds in the US population.

Reducing the bias of non-probability sample estimators through inverse probability weighting with an application to Statistics Canada's crowdsourcing data

Jean-François Beaumont

Non-probability samples are being increasingly explored in National Statistical Offices as an alternative to probability samples. However, it is well known that the use of a non-probability sample alone may produce estimates with significant bias due to the unknown nature of the underlying selection mechanism. To reduce this bias, data from a non-probability sample can be integrated with data from a probability sample that contains auxiliary variables in common with the non-probability sample. We focus on inverse probability weighting methods, which involve modelling the probability of participation in the non-probability sample. First, we consider the logistic model along with the pseudo maximum likelihood method of Chen, Li and Wu (2020). We propose a variable selection procedure based on a modified Akaike Information Criterion (AIC) that properly accounts for the data structure and the probability sampling design. We also propose a simple rank-based method of forming homogeneous post-strata. Then, we extend the Classification and Regression Trees (CART) algorithm to this data integration scenario, while again properly accounting for the probability sampling design. A bootstrap variance estimator is proposed that reflects two sources of variability: the probability sampling design and the participation model. Our methods are illustrated using Statistics Canada's crowdsourcing and survey data.

Program

United States Department of Agriculture
Tuesday, March 1, 2022

Opening Remarks
Linda J Young (*USDA NASS*)

Program Chair
Brian Harris-Kotejin

Keynote Speakers
Courtney Kennedy
Yan Li
Jean-François Beaumont

Moderator
Brian Harris-Kotejin

Hansen Lecture Committee
Brian Harris Kojetin (Chair)
Steve Cohen (Past Chair)
Partha Lahiri (Chair-Elect)
Keith Rust (Westat Representative)
Katherine J. Thompson
Linda J. Young (NASS Representative)

29th Morris Hansen Lecture Series



Morris Hansen Memorial Lecture Series

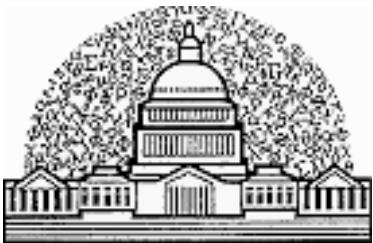
Morris Howard Hansen has been described as the most influential statistician in the development of survey methodology in the twentieth century. Early in his Census Bureau career he put together a staff to define the principal problems in the conduct of surveys, investigate these problems, and develop statistical methods to address them. Morris and his staff then widely distributed the results of their efforts, thus influencing statistical agencies all over the world. Generations of statistical students have learned from and been influenced by *Sample Survey Methods and Theory*,

Morris was also known as an innovator and a leader in adapting electronic tools, such as computers and mark-reading sensors, to statistics. After his outstanding Census Bureau career, Morris joined Westat which was at the time a fairly small statistical research company. Morris again assembled a strong staff and expanded Westat's scope to take on large federal government statistical problems.

Morris also made outstanding contributions to professional organizations, serving as the president of both the American Statistical Association and the Institute of Mathematical Statistics and as the first president of the International Association of Survey Statisticians. He was elected to the National Academy of Sciences in 1976 and was an important member of many Academy committees and panels.

There have been many tributes to Morris since his death in 1990, such as memorial issues of both the *Journal of Official Statistics* and *Survey Methodology*. Westat issued a grant to the Washington Statistical Society to honor Morris with an annual lecture series. The series has been so successful in attracting top quality presentations on a wide variety of topics—in keeping with Morris' broad interests—that Westat has added to the original grant.

Sponsored by:



Washington Statistical Society

