WASHINGTON STATISTICAL SOCIETY

Gertrude Cox Award Lecture

Sponsored by the Washington Statistical Society and RTI International

Design-based thinking in the model-based practice:
Shaping survey methodology for modern data science

Dr. Yajuan Si
Research Associate Professor, Survey Research Center
Associate Director, Michigan Program in Survey and Data Science
University of Michigan

Date & Time: Wednesday, June 3, 2026, 12:00–1:00 PM ET

Location: Online via RTI Zoom

Join Zoom: https://rtiorg.zoom.us/j/5793514370?pwd=VjdHTWpCd012SGs2ZEJ2TEZTbEJUZz09&omn=98281746158&from=addon

Meeting ID: 579 351 4370 | **Passcode:** 676101

Abstract

Survey methodology faces a paradox: just as probability sampling, error quantification, and selection bias adjustment are most needed, amid declining response rates, rising operational costs, and growing reliance on found data, the field is increasingly perceived as less central than model-based or algorithmic analytics. In this lecture, I argue that design-based reasoning is not a legacy constraint but a practical discipline for making modern data science more reliable, transparent, and decision-relevant.

I present lessons from my work that connects survey design and inference to contemporary model-based practice. First, I present multilevel regression and poststratification as a unifying framework for integrating surveys with auxiliary information to calibrate small area

estimation. Second, drawing on population-science applications in large-scale cohort studies, I discuss how reference probability samples and carefully chosen auxiliary variables can improve population generalizability and also inform what data should be collected next. Third, I discuss hybrid data collection and inference that combine probability and nonprobability samples.

These examples highlight practical design questions: Which auxiliary variables are needed? When is probability sampling essential, and when can nonprobability data safely add efficacy? Which assumptions are empirically testable through design-informed diagnostics? Using these examples at the intersection of survey design, causal effect generalizability, data integration, and reproducible statistical computing, I conclude that the most useful models in practice are those that respect the design.

Biography

Dr. Yajuan Si is a Research Associate Professor at the Survey Research Center and Associate Director of the Michigan Program in Survey and Data Science at the University of Michigan. She joined Michigan in 2017 after three years on the faculty at the University of Wisconsin–Madison in biostatistics and population health. She earned her Ph.D. in Statistical Science from Duke University and completed postdoctoral training at Columbia University.

Her research develops statistical methods that connect data analysis to study design, with an emphasis on Bayesian approaches to population inference. Her work spans design- and model-based survey inference, data integration, missing data methods, confidentiality protection, and causal inference, with applications in the social and health sciences. A central focus of her research program is improving population inference from complex and integrated data sources by addressing selection and nonresponse bias through principled modeling and design-informed adjustment. She is the recipient of the 2026 Gertrude Cox Award.