All Times ET	Schedule
8:00 am	Registration and Coffee/Light Breakfast
8:50 am	Welcome: *Andrea Troxel, Professor and Director of Biostatistics, NYU Langone*
9:00 – 10:00 am	**Keynote Address: *"Integrating Data for Causal Inference"*** Elizabeth Stuart, Professor and Chair of Biostatistics, Johns Hopkins Bloomberg School of Public Health
10:00 – 10:30 am	Coffee Break
10:30 – 11:30 am	**Session 1: Causal Inference for Real World Data** *Moderator: Michele Santacatterina, Assistant Professor of Biostatistics, NYU Langone* José Ramon Zubizarreta, Professor, Harvard University Xiao Wu, Assistant Professor, Columbia University Natalie Levy, Director of Scientific Strategy, Aetion
11:30 am – 1:00 pm	Lunch
1:00 – 2:00 pm	**Session 2: Methodological Developments for Causal Inference under Complex Data Structures** *Moderator: Iván Díaz, Associate Professor of Biostatistics, NYU Langone* Matteo Bonvini, Assistant Professor, Rutgers University Youjin Lee, Assistant Professor, Brown University Ashkan Ertefaie, Associate Professor, Rochester University
2:00 – 2:30 pm	Coffee Break
2:30 – 3:30 pm	**Session 3: Bayesian Approach to Causal Inference** *Moderator: Samrachana Adhikari, Associate Professor of Biostatistics, NYU Langone* Michael Harhay, Assistant Professor, University of Pennsylvania Arman Oganisian, Assistant Professor, Brown University Laura Forastiere, Associate Professor, Yale University
3:30 – 4:00 pm	Closing Remarks: *Hayley Belli, Assistant Professor of Biostatistics, Chair of Organizing Committee, NYU Langone*
4:00 – 5:00 pm	Reception

NYU Langone Health

Biostatistics Division
Department of Population Health

2024 Biostatistics Symposium

Advances in Causal Inference Methods

Friday, October 25, 2024
8:00 am – 5:00 pm ET

180 Madison Avenue, 2nd Floor
New York, NY 10016

NYU Langone Health

Speaker Titles, Abstracts, and Biographies:

2024 NYU Langone Biostatistics Symposium: Advances in Causal Inference
Friday, October 25, 2024

Keynote: 9:00 – 10:00am

Elizabeth Stuart:

Title: Integrating data for causal inference

Abstract:
Many causal questions of interest cannot be answered through analysis of a single dataset, and as data becomes increasingly available, there is more and more interest in leveraging that data to answer nuanced questions. Such questions might include examining the generalizability of randomized trial results to target populations, to better understanding of effect heterogeneity by combining small (unbiased) randomized trials with large (but confounded) non-experimental data sources. This talk will discuss methods for causal inference in such integrated datasets, including both the promise and potential for doing so, as well as implementation challenges, such as when the measures in the different data sources are discordant. Motivating examples will come from medicine and public health, and with lessons for a range of fields, and with final comments on the broader field of evidence synthesis for causal inference.

Bio: Elizabeth A. Stuart, Ph.D. is the Frank Hurley and Catharine Dorrier Chair and Bloomberg Professor of American Health in the Department of Biostatistics at the Johns Hopkins Bloomberg School of Public Health, with joint appointments in the Department of Mental Health and the Department of Health Policy and Management. A statistician by training, her research interests are in design and analysis approaches for estimating causal effects in experimental and non-experimental studies, including questions around the external validity of randomized trials and the internal validity of non-experimental studies. She has published over 350 papers and has received research funding for her work from the National Science Foundation, the Institute of Education Sciences, the WT Grant Foundation, and the National Institutes of Health and has served on advisory panels for the National Academy of Sciences, the US Department of Education, and the Patient Centered Outcomes Research Institute. She currently serves on the NASEM Committee on National Statistics and co-chairs NASEM's Committee on Applied and Theoretical Statistics.

Session 1: Causal Inference for Real World Data, 10:30 – 11:30am

José Ramon Zubizarreta:

Title: An Anatomy of Event Studies: Hypothetical Experiments, Exact Decomposition, and Weighting Diagnostics

Abstract: In recent decades, event studies have emerged as a central methodology in health and social research for evaluating the causal effects of staggered interventions. In this paper, we analyze event studies from the perspective of experimental design and focus on the use of information across units and time periods in the construction of effect estimators. As a particular case of this approach, we offer a novel decomposition of the classical dynamic two-way fixed

effects (TWFE) regression estimator for event studies. Our decomposition is expressed in closed form and reveals in finite samples the hypothetical experiment that TWFE regression adjustments approximate. This decomposition offers insights into how standard regression estimators borrow information across different units and times, clarifying and supplementing the notion of forbidden comparison noted in the literature. We propose a robust weighting approach for estimation in event studies, which allows investigators to progressively build larger valid weighted contrasts by leveraging increasingly stronger assumptions on the treatment assignment and potential outcomes mechanisms. This weighting approach also allows for the generalization of treatment effect estimates to a target population. We provide diagnostics and visualization tools and illustrate these methods in a case study of the impact of divorce reforms on female suicide.

Bio: José R. Zubizarreta, PhD, is Professor in the Department of Health Care Policy at Harvard Medical School, Professor in the Department of Biostatistics at Harvard School of Public Health, and Faculty Affiliate in the Department of Statistics at the Faculty of Arts and Sciences at Harvard University. His work centers on the development of statistical methods for causal inference and impact evaluation to advance research in health care and public policy. He was a Fulbright Scholar and has received funding from the Alfred P. Sloan Foundation. He serves as Associate Editor of the Annals of Applied Statistics, Biometrics, and the Journal of Royal Statistical Society. He is a Fellow of the American Statistical Association, and is a recipient of the Kenneth Rothman Award, the William Cochran Award, and the Tom Ten Have Memorial Award.

Xiao Wu:

Title: Synthetic control via covariate balancing with applications in long-term health impacts of extreme climate events

Abstract: Hurricanes and tropical storms, collectively called tropical cyclones (TCs), are fast-spinning storm systems with a low-pressure center, strong winds, and spiraling thunderstorms. TCs can cause extensive infrastructure damage due to their strong winds, torrential rain, and storm surges, and a major threat to health and welfare in the US, with TC-related costs over $1.3 trillion in the past four decades. Currently, there are critical knowledge gaps in understanding multi-year trends of health conditions post-TC exposure in affected communities. The lack of suitable statistical methods is a significant obstacle in investigating long-term health effects of TCs. Recently, there has been growing interest in using Synthetic Control (SC) methods for evaluating long-term effects of various public health interventions. Traditional SC approaches were developed for relatively small panel data, rendering them ill-suited for large-scale environmental data. Additionally, traditional SC methods typically adjust for only the historical outcome trajectory and, at most, a limited number of covariate paths of a single exposed unit. In contrast, TC studies often gather an extensive list of pre-exposure covariates that describe the characteristics of each disaggregated-level spatial unit, which must be considered when creating synthetic control units. To address these gaps, we extend the SC methods with tailored balance conditions, which can be approached as a covariate balancing problem. This problem can be solved using an unconstrained convex optimization formulation proposed for high-dimensional propensity score estimation. Using the SC methods, we estimate the causal effects of annual county-level TC exposure on the social vulnerability index (SVI) between 2010 and 2020 based on all TCs that took place in the US. Our findings indicate that TC exposures generally adversely impact SVI immediately after exposure, while exposed regions recover from long-term TC disasters.

Bio: Xiao Wu is an Assistant Professor of Biostatistics at Columbia University, and a member of Columbia Data Science Institute. His research expertise lies in developing statistical and causal inference methods to address methodological needs in environmental health research. The key milestone of his research is to provide scientific evidence and policy solutions to mitigate the adverse impacts of environmental factors in a rapidly changing climate.

Natalie Levy:

Title: Transforming Real-World Data to Real-World Evidence: Causal Inference for Non-Interventional Studies to Inform Regulatory Decision-Making

Abstract: Regulators of medical products are increasingly recognizing the value of real-world data (RWD) - data generated in the course of routine healthcare - to inform decision-making about the effectiveness and safety of drugs, biologics, and medical devices. To produce real-world evidence (RWE), non-interventional studies using RWD must meet equivalent standards for causal inference as randomized controlled trials. The use of RWD presents unique challenges to causal inference, including specific sources of bias arising from the nature of the data itself and difficulties defining Time Zero using RWD. In this session, we will discuss these issues and review available tools, such as the hypothetical target trial approach and SPACE/SPIFD2 framework, to aid in the design and conduct of valid and transparent RWE studies.

Bio: Dr. Natalie Levy is a Director of Scientific Strategy at Aetion. In this role, she provides methodological and analytic support to teams working with pharmaceutical companies across a variety of therapeutic areas, contributes to research to inform RWE best practices, and is a scientific lead for the Aetion Coalition to Advance Real-World Evidence through Randomized Controlled Trial Emulation (CARE) Initiative. Natalie completed her doctoral training in Epidemiology at Columbia University, with a focus in epidemiologic methods and causal inference. Previously, she worked at the New York City Department of Health and Mental Hygiene where she managed the Bureau of Tuberculosis Control's electronic surveillance registry and developed and maintained data standards for epidemiologic research, analysis, and reports. Natalie has designed and taught courses in epidemiologic study design and analysis of complex surveys and has published on a variety of topics including disease transmission, substance use, health policy evaluation, and causal inference.

Session 2: Methodological Developments for Causal Inference under Complex Data Structures, 1:00–2:00pm

Matteo Bonvini:

Title: Time-varying Heterogeneous Effects Estimation

Abstract: We study the problem of estimating heterogeneous treatment effects across multiple time points under no-unmeasured-confounding. Following the principles of DR-Learning, we aim to design a procedure to estimate the conditional mean of the potential outcomes viewed as a function of the baseline covariates and the whole treatment sequence, where the treatment is possibly multi-valued. For example, in the case of a large number of time points, the procedure would be able to regularize or smooth across treatment trajectories to facilitate inference. We also aim to develop bootstrap-based uniform confidence bands that cover the true effect uniformly across covariate values, treatments, and time. Finally, we plan to apply our methods

to an observational study designed to investigate the effects of time-varying combinations of surgery, chemotherapy, and radiation therapy on overall survival for patients affected by certain gynecologic cancers.

Bio: Matteo Bonvini is an Assistant Professor in the Department of Statistics at Rutgers, the State University of New Jersey. He obtained a Ph.D. in Statistics from Carnegie Mellon University in 2023, advised by Professor Edward H. Kennedy. His thesis was awarded the department's Umesh K. Gavaskar Best Dissertation Award. Before starting the Ph.D., he worked as Analyst at Cornerstone Research and graduated with a B.A. in Statistics from Harvard University in 2016. His current research focuses on developing methodology at the intersection of nonparametric statistics and causal inference.

Youjin Lee:

Title: Replicable causal research with instrumental variables

Abstract: In observational studies, unmeasured confounders can introduce bias in causal estimates, and this bias is often systematic and recurs in replicated studies. Instrumental variables (IVs) have been widely used to estimate the causal effect of a treatment on an outcome in the presence of unmeasured confounders. When several IVs are available and the instruments are subject to biases that do not completely overlap, a careful analysis using these instruments can produce orthogonal pieces of evidence (i.e., evidence factors) that, when combined, strengthen causal conclusions while avoiding systematic bias. In this talk, I will introduce a novel approach that combines multiple evidence factors with multiple IVs that are causally ordered and potentially influenced by different levels of decision-makers. We apply our proposed approach to evaluate the effect of extended time accommodations.

Bio: Youjin Lee is a Manning Assistant Professor of Biostatistics at Brown University. Her research focuses on the development of statistical and causal inference methods that address unmeasured confounding and dependent observations. Dr. Lee is currently interested in applications to policy effect evaluation and effective connectivity studies. She also serves as an Associate Editor for Reproducibility for the Journal of the American Statistical Association and as Secretary for the Society for Causal Inference.

Ashkan Ertefaie:

Title: Causal Mediation Analysis: Selection with Asymptotically Valid Inference

Abstract: Researchers are often interested in learning not only the effect of treatments on outcomes, but also the mechanisms that transmit these effects. A mediator is a variable that is affected by treatment and subsequently affects outcome. Existing methods for penalized mediation analyses may lead to ignoring important mediators and either assume that finite-dimensional linear models are sufficient to remove confounding bias, or perform no confounding control at all. In practice, these assumptions may not hold. We propose a method that considers the confounding functions as nuisance parameters to be estimated using data-adaptive methods. We then use a novel regularization method applied to this objective function to identify a set of important mediators. We consider natural direct and indirect effects as our target parameters. We then proceed to derive the asymptotic properties of our estimators and establish the oracle property under specific assumptions. Asymptotic results are also presented in a local setting which contrast the proposal with the standard adaptive lasso.

We also propose a perturbation bootstrap technique to provide asymptotically valid post-selection inference for the mediated effects of interest.

Bio: Ashkan Ertefaie is an Associate Professor of Biostatistics and Neurology in the Department of Biostatistics and Computational Biology at the University of Rochester. His methodological research interests lie in causal inference, dynamic treatment regimes, sequential multiple assignment randomized trials, comparative effectiveness studies using electronic health records, instrumental variable analyses, high-dimensional data analysis, post selection inference, and survival analysis. Particularly, he is interested in methodologies that lead to constructing the optimal sequence of treatments in chronic diseases where patients are monitored and treated throughout their life. Thus far, he has been primarily motivated by applications in health services, medicine and criminal justice.

Session 3: Bayesian Approach to Causal Inference, 2:30–3:30pm

Michael Harhay:

Title: A Bayesian machine learning approach for estimating heterogeneous survivor causal effects: Applications to a critical care trial

Abstract:
Assessing heterogeneity in the effects of treatments has become increasingly popular in the field of causal inference and carries important implications for clinical decision-making. While extensive literature exists for studying treatment effect heterogeneity when outcomes are fully observed, there has been limited development in tools for estimating heterogeneous causal effects when patient-centered outcomes are truncated by a terminal event, such as death. Due to mortality occurring during study follow-up, the outcomes of interest are unobservable, undefined, or not fully observed for many participants in which case principal stratification is an appealing framework to draw valid causal conclusions. Motivated by the Acute Respiratory Distress Syndrome Network (ARDSNetwork) ARDS respiratory management (ARMA) trial, we developed a flexible Bayesian machine learning approach to estimate the average causal effect and heterogeneous causal effects among the always-survivors stratum when clinical outcomes are subject to truncation. We adopted Bayesian additive regression trees (BART) to flexibly specify separate mean models for the potential outcomes and latent stratum membership. In the analysis of the ARMA trial, we found that the low tidal volume treatment had an overall benefit for participants sustaining acute lung injuries on the outcome of time to returning home but substantial heterogeneity in treatment effects among the always-survivors, driven most strongly by biologic sex and the alveolar-arterial oxygen gradient at baseline (a physiologic measure of lung function and degree of hypoxemia). These findings illustrate how the proposed methodology could guide the prognostic enrichment of future trials in the field.

Bio: Dr. Michael Harhay is an Assistant Professor of Epidemiology, Medicine, and Statistics and Data Science at the University of Pennsylvania and an Honorary Research Fellow at the MRC Clinical Trials Unit at University College London. His research program is supported by multiple grants from the NIH and PCORI and spans clinical trials, large epidemiologic studies, and quality improvement initiatives within health systems and among seriously and critically ill patients. He is involved in a wide range of international research activities, including randomized trial data monitoring boards, steering committees, and research consortia, and currently serves as deputy editor of the American Journal of Respiratory and Critical Care Medicine and is an editor of the International Journal of Epidemiology.

Arman Oganisian:

Title: Bayesian Semiparametric Inference for Sequential Decision-Making with Incomplete Information

Abstract: Many chronic diseases are managed over time via a sequence of treatment decisions. While time-varying covariates are used to tailor the treatment, sporadic monitoring leaves us with incomplete covariate information. Additionally, censoring leaves us with incomplete outcome information. In our primary data example, patients with pediatric acute myeloid leukemia (AML) move through a sequence of treatment courses where a decision is made to withhold a scheduled course of anthracycline chemotherapy (ACT). Since ACT is cardiotoxic, echocardiograms are sometimes - but not always - conducted to inform the withholding decision with the goal of improving survival. We frame this as a dynamic treatment rule (DTR) problem and identify potential marginal survival rates under different DTRs for making monitoring and withholding decisions jointly. Bayesian semiparametric models are used to model continuous-time transitions between treatment courses (recurrent states) and death (the absorbing state) conditional on monitoring/withholding decisions. A g-computation procedure simulates the transition process to produce posterior draws of the marginal survival rates.

Bio: Arman Oganisian is an Assistant Professor of Biostatistics at Brown University's School of Public Health. His research is focused on developing Bayesian nonparametric and machine learning methods for causal inference. More information can be found on his web site https://stablemarkets.netlify.app/

Laura Forastiere:

Title: causal inference on partially observed epidemic network with application to social program on behaviour change in rural Honduras

Abstract: In this work, we study how behavioral interventions affect the prevalence of behaviors which can diffuse among the connected units per cluster with some probability and a varying time. We consider the setting where we can observe the set of treated units, the network structure and the behavior prevalence only at baseline and at an end time T. To investigate the problem, we formulate a exposure mapping under the baseline conditions of epidemic network models (e.g., SIS and SI models governed by certain parameters). In order to estimate the causal effect between any two exposures, we utilize a MCMC framework for estimating the parameters of epidemic models, with a data augmenting procedure to impute the missing processes between the baseline and the endline. We then construct IPW estimators based on these parameters and the probabilities of exposures. We utilize these estimators to evaluate the causal effects with different treatment rates and strategies, baseline behavior prevalence as well as network structures. We also apply the proposed procedures and estimators among different villages of Honduras, to evaluating the change of adoption rates on certain behaviors based on social programs.

Bio: Laura Forastiere is an Associate Professor in the Department of Biostatistics at Yale School of Public Health. Her methodological research is focused on methods for assessing causal inference for evidence-based research, exploring the mechanisms underlying the effect of an intervention including causal pathways through intermediate variables or mechanisms of peer

influence and spillover between connected units. Her research explores modeling, inferential, and other methodological issues that often arise in applied problems with complex clustered and network data, and standard statistical theory and methods are no longer adequate to support the goals of the analysis. Laura is eager to apply advanced statistical methodology to provide evidence on effective strategies to improve the health and wellbeing of vulnerable populations. She is particularly interested in exploring behavioral interventions that, relying on theories of behavioral economics and social phycology, exploit social interactions and peer influence among individuals. She is involved in many program evaluations and research studies in low- and middle-income countries on malaria, HIV and other STDs, maternal and child health, nutrition, cognitive development, health insurance and microcredit.