BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Liangyuan Hu

eRA COMMONS USER NAME (credential, e.g., agency login): LIANGYUAN

POSITION TITLE: Associate Professor of Biostatistics and Epidemiology

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	Completion Date MM/YYYY	FIELD OF STUDY
South China University of Technology, China	BEng	07/2002	Civil Engineering
University of Alberta, Canada	MSc	04/2007	Statistics
Brown University, USA	PhD	06/2015	Biostatistics

A. Personal Statement

I am currently an associate professor at Rutgers Univerisy School of Public Health. My research concerns the development of statistical methods for causal inference, missing data, machine learning and longitudinal data analysis, with applications spanning a wide range of disease areas such as cognitive impairment, cancer, COVID-19, and cardiovascular diseases.

My recent publications on statistical methodology include the development of Bayesian machine learning methods for causal inference, longitudinal data analysis, and variable selection. As part of my collaborations in biomedical research, I took the lead with the design of and implementation of the cutting-edge statistical methods for several projects, which led to publications in high-impact journals. In addition to my publications on methodology and clinical applications, I am coauthor of the book chapter, *A Bayesian Perspective on Assessing Sensitivity to Assumptions about Unobserved Data* in: *Handbook of Missing Data Methodology*.

As the PI of multiple methodology grants, I am actively involved in the development of statistical and machine learning methods to study the comparative effects of complext interventions using large-scale health datasets. Addtionaly, I serve as co-I on multiple grants, contributing my statistical expertise to the development of novel methodologies for stepped wedge cluster randomized trials and the design and analysis of studies with cognitive outcomes.

Specific work of direct relevance to this project includes the following:

1. R01 HL159077
   NIH/NHLBI
   Bayesian machine learning for causal inference with incomplete longitudinal covariates and censored survival outcomes
   Liangyuan Hu (PI)
   05/10/2022-4/30/2027

2. ME-2021C2-23685
   PCORI
   Robust longitudinal causal inference methods with machine learning
   Liangyuan Hu (PI)

8/1/2022 – 7/31/2025

3. R21 CA245855
   NIH/NCI
   Flexible Bayesian approaches to causal inference with multilevel survival data and multiple treatments
   Liangyuan Hu (PI)
   07/16/2020-06/30/2023

4. R01AG078211
   NIH/NIA
   Determinants of Individual Differences in the Efficacy of Aerobic Exercise to Improve Brain Health and Reduce Alzheimer Disease Risk in Older African Americans (PI: Mark Gluck)
   Liangyuan Hu (Co-I)
   9/15/2022 – 8/31/2027

5. TBA
   PCORI
   Towards Improved Design and Analysis of Stepped Wedge Trials by Leveraging Baseline Information (PI: Fan Li)
   Liangyuan Hu (SubPI)
   9/1/2023 – 8/31/2026

B. Positions, Scientific Appointments, and Honors

Positions and Scientific Appointments

2021 –	Associate Professor, Rutgers University, USA
2021 –	Editorial Board Member, Journal of Clinical Oncology
2021	External methodology reviewer, Patient-Centered Outcomes Research Institute
2020	Reviewer (ad hoc), NIH Study Sections, Clinical Oncology (CONC)
2019	External methodology reviewer, Patient-Centered Outcomes Research Institute
2018	JSM organizer, Invited Paper “Modern Statistical Methods in Comparative Effectiveness Research”
2018	Grant reviewer, The Medical Research Council, UK Research and Innovation
2017	JSM organizer, Invited Paper “New Developments in Sensitivity Analysis for Unmeasured Confounding”
2016	JSM organizer, Invited Paper “Data-driven Modeling in Medical & Health Policy Decision Making”
2016 – 2021	Assistant Professor, Icahn School of Medicine at Mount Sinai, USA
2015 – 2016	Associate, Analysis Group, Inc., USA
2013 – 2014	Organizing Committee Member & Co-chair, Atlantic Causal Inference Conference
2008 – 2009	Management Science Analyst, Edmonton Police Service, Canada
2007 – 2008	Statistician, Alberta Health & Wellness, Canada

Awards and Honors

2021	Delta Omega Honorary Society in Public Health
2019	American Statistical Association Outstanding Statistical Application Award, Denver, CO.
2017	Young Investigator Travel Award, MCW Biostatistics Conference, Milwaukee, WI.
2015	Outstanding Paper Award, Health Policy Statistics Section, American Statistical Association
2015	Best Graduate or Post-Doctoral Trainee Poster Presentation, SPH, Brown University
2015	CFAR Trainee Support Award, Lifespan/Turfs/Brown Center for AIDS Research
2014	Best Poster Award, 18th International Workshop on HIV Observational Databases, Spain
2007	Josephine M. Mitchell Scholarship for Excellence in Academic Achievements, University of Alberta
2006	Canadian Federation of University Women — Edmonton Academic Awards Fund, Canada

C. Contributions to Science

1. **Causal inference and machine learning**. While large healthcare data sets contain rich information, complications associated with these data such as selection bias and multilevel treatments must be addressed by rigorous causal inference methods in order to draw conclusions about treatment effects. I and colleagues developed statistical and machine learning methods for causal inference with multiple treatments and a binary outcome (Hu et al. 2020), and with multilevel treatments and rare outcomes (Hu and Gu, 2021), and for sensitivity analysis for baseline unmeasured confounding (Hu et al. 2022). I provided insights into the role of the Bayesian machine learning technique, Bayesian Additive Regression Trees, in causal inference (Hu, 2020). I and colleagues developed ways in which we can use modern machine/deep learning methods to estimate heterogeneous causal effects from observational censored survival data (Hu et al. 2021).

a. **Hu L**, Zou J, Gu C, Ji J, Lopez M, Kale M (2022). A flexible sensitivity analysis approach for unmeasured confounding with multiple treatments and a binary outcome with application to SEER-Medicare lung cancer data. *Annals of Applied Statistics* 16(2): 1014-1037. PMCID: PMC9835106
b. **Hu L**, Ji J, Li F (2021). Estimating heterogeneous survival treatment effect in observational data using machine learning. *Statistics in Medicine* 40(21): 4691-4713. PMID: 34114252 PMCID: PMC9827499
c. **Hu L**, Gu C (2021). Estimation of causal effects of multiple treatments in healthcare database studies with rare outcomes. *Health Services & Outcomes Research Methodology* 21: 287-308.
d. **Hu L**, Gu C, Lopez M, Ji J , Wisnivesky J (2020). Estimation of Causal Effects of Multiple Treatments in Observational Studies. *Statistical Methods in Medical Research* 29(11): 3218-3234. PMCID: PMC7534201

2. **Longitudinal studies**. The censored survival outcomes and complex longitudinal data structures in patient-oriented outcomes research require specialized methodology. I led the development of statistical methods to tackle the difficult problem of drawing causal effects about a time-varying exposure on survival outcomes when both the exposure and the survival time may be censored (Hu et al. 2018), and for conducting comparative effectiveness analysis of dynamic treatment regimens when the longitudinal outcome and death are sparse with various missing data patterns (Hu and Hogan, 2019). I develop statistical theory and methods for propensity score weighting analysis of survival outcomes using pseudo-observations (Zeng et al. 2022). More recently, I and colleagues developed a joint marginal structural proportional hazards model and novel weighting schemes in continuous time to evaluate the effectiveness of multiple intermittent COVID-19 treatments (Hu et al. 2022).

a. **Hu L**, Hogan JW (2019). Causal comparative effectiveness analysis of dynamic continuous-time treatment initiation rules with sparsely measured outcome and death. *Biometrics*; 75(2): 695-707. PMID: 30638268 PMCID: PMC9831746
b. **Hu L**, Hogan JW, Mwangi AM, Siika A (2018). Modeling the causal effect of treatment initiation time on survival: Application to HIV/TB coinfection. *Biometrics;* 74(2): 703-713. PMCID: PMC5874166
c. Zeng S, Li F, **Hu L,** Li F (2022). Propensity score weighting analysis of survival outcomes using pseudo-observations. *Statistica Sinica;* Published Online.
d. **Hu L**, Li F, Ji J, Joshi H, Scott E (2022). Estimating the causal effects of multiple intermittent treatments with application to COVID-19. *arXiv preprint*. arXiv:2109.13368v3. PMCID: PMC8722604

3. **Identification of risk factors and missing data.** The missing data issue is ubiquitous in health investigations. In a recent publication (Hu et al. 2021), I developed a general approach to effectively select useful predicting variables when both covariates and outcomes have missing data by combining bootstrap imputation and machine learning modeling techniques. We further developed an inference-based approach that offers substantial computational savings for effective variable selection using likelihood-based Bayesian machine learning (Lin et al. 2021). In a chapter of the book *Handbook of Missing Data Methodology* I coauthored with Drs Hogan and Daniels, we provided a Bayesian perspective on assessing sensitivity to assumptions about unobserved data (Hogan et al. 2014).

a. **Hu L**, Lin JY*, Ji J (2021). Variable selection with missing data in both covariates and outcomes: Imputation and machine learning. *Statistical Methods in Medical Research;* 30(12): 2651-2671. PMID: 34696650

b. Lin JY*, Hu L, Huang C, Lawrence S, Govindarajulu U (2021). Strategies for variable selection in large-scale database studies with missing covariates and outcome data. *BMC Medical Research Methodology* 22: 132. PMCID: PMC9066834
c. Hogan JW, Daniels MJ, **Hu L** (2014). A Bayesian perspective on assessing sensitivity to assumptions about unobserved data. In: *Handbook of Missing Data Methodology*. CRC Press. Chap.18:405-434.

4. **Novel statistical application of machine learning.** I led applications of cutting-edge statistical machine learning approaches to large healthcare data for solving several important domain area research questions, including cardiovascular disease (Hu et al. 2021, Hu et al. 2020a), cancer (Hu et al. 2020b) and inflammatory bowel disease (Ungaro et al. 2021). These works demonstrate my interest and ability to engage with clinical researchers and promote improving patient-oriented outcomes with methodological rigor.

a. **Hu L**, Li Y (2022). Using tree-based machine learning for health studies: Literature review and case series. International Journal of Environmental Research and Public Health; 19(23): 16080. PMCID: PMC9736500
b. **Hu L**, Liu B, Ji J, Li Y(2020a). Tree-based machine learning to identify and understand major determinants for stroke at the neighborhood level. *Journal of American Heart Association*; 9(22): e016745. PMCID: PMC7763737
c. **Hu L**, Li L, Ji J, Sanderson M (2020b) Identifying and understanding determinants of high healthcare costs for breast cancer: a quantile regression machine learning approach. *BMC Health Services Research*; 20: 1066. PMCID: PMC7684910
d. Ungaro R, **Hu L**, Ji J, Nayar S, Kugathasan S, Denson L, Hyams J, Dubinsky M, Sands B, Cho J (2021). Machine learning identifies novel blood protein predictors of penetrating and stricturing complications in newly diagnosed paediatric Crohn's disease. *Alimentary Pharmacology & Therapeutics*; 53(2): 281-290. PMCID: PMC7770008

Complete List of Published Work in MyBibliography:

https://www.ncbi.nlm.nih.gov/myncbi/1hM9F5OG5YK51/bibliography/public/