

Jie Kate Hu, PhD

Curriculum Vitae

Department of Biostatistics
655 Huntington Avenue
Boston, MA 02115

18 Whitney Ave, Unit 5
Cambridge, MA, 02139

Email: khu@hsph.harvard.edu

LinkedIn: www.linkedin.com/in/kate-hu

Website: www.katehu.com

EDUCATION

- Ph.D. Department of Biostatistics, University of Washington, Seattle, 2014
Dissertation: [*A Z-estimation System for Two-phase Sampling with Applications to Additive Hazards Models and Epidemiologic Studies*](#)
Committee: Norman Breslow (Co-Chair), Gary Chan (Co-Chair), Jon Wellner, Peter Gilbert, Ying Chen, Loveday Conquest
- M.S. Department of Biostatistics, Harvard University, 2008
- B.S. Department of Biochemistry (*First Class Highest Honors*), University of Hong Kong, 2006

Exchange Study

University of California, Los Angeles, 2005-2006

Peking University, China 2002-2003

PROFESSIONAL EXPERIENCE

Academic

2022-present National Research Service Award Postdoctoral Fellow, Harvard T.H. Chan School of Public Health, Mentors: Francesca Dominici and Eric Tchetgen Tchetgen

Industry

2019-2021 Head of Data Science/Principal Data Scientist, [Aclima Inc](#), San Francisco

- Led research and software development on scalable sampling designs for mapping street-by-street air quality and greenhouse emission with mobile sensing platforms, deployed in multiple metropolitan areas
- Led research on scalable bias adjustment algorithms and development of hyper-local air quality maps, serving multiple regulatory agencies and environmental justice communities

- Built data science team from scratch and forged collaborations across academia, governments, and industry

2014–2019 Senior Quantitative Researcher, [The Climate Corporation](#), (Bayer Crop Science), San Francisco

- Developed sampling methods, patents, and experimental design software for model calibration, validation, sensor placement, and treatment allocation used repeatedly by scientists to collect field data
- Developed analysis methods to 1) repurpose trial data for new hypotheses testing and 2) estimate average treatment effect with imperfect trial data
- Secured budget and implemented large-scale field trials to test multiple precision treatment algorithms over five states in Midwest
- Developed location-specific and weather-adaptive nitrogen treatment algorithms to help farmers maximize crop yield amid climate change

2008–2014 Research Assistant, Fred Hutchinson Cancer Research Center, Seattle

- Developed semi-parametric hazards models with applications to a HIV/AIDS study
- Investigated gene-gene interaction tests for Genome Wide Association Studies
- Evaluated prediction capability of biomarkers and examined biases of ROC curve

2012–2014 Consultant, School of Medicine, University of Washington, Seattle

- Consultation for the design and analysis of pharmaco-epidemiology studies

FELLOWSHIPS & AWARDS

Fellowship

2022-2024 NIH National Research Service Award Postdoctoral Fellowship (\$130,000)
 2007-2008 The Jeffries Wyman Scholarship, Harvard University (\$30,000)
 2003-2006 Hong Kong Government Fellowship, University of Hong Kong (\$45,000)

Awards

2013 American Statistical Association to Joint Statistical Meetings
 2013 Graybill Conference for Modern Survey Statistics
 2012 National Science Foundation to World Congress on Probability and Statistics, Istanbul, Turkey

Leadership Award

2021 [No.1 of the Most Innovative Companies in Data Science](#), Role: Built the Data Science team and Head of Data Science

GRANT APPLICATION

under review U.S. Forest Service, “Adaptive Sampling Methods and System for Estimating Forest Status and Trends with Reduced Costs”

2022 Google’s AI for Social Good, “Using AI to drive intelligent land reuse: decision support tools to help direct 2.1 billion USD in contaminated site redevelopment investment,” not funded

2022 The Harvard Chan - National Institute of Environmental Health Sciences Center for Environmental Health Pilot Project, “Development and Dissemination of Negative Controls to Adjust for Unmeasured Confounding: Untangle the Effects of Wildfire Smoke and Other Air Pollutants on Health Outcomes,” not funded

PATENTS

pending **Hu, J.** System and Method for Modular Design of Statistical Models. (U.S. Application Serial No. 18/051,403, pending)

2023 **Hu, J,** Jerkins, J, Goebel, N. [Routing method for mobile sensor platforms](#). (U.S. Patent No. #11,553,318)

2022 **Hu, J** & Carrion C., [Using Causal Learning Algorithms to Assist in Agricultural Management Decisions](#). (U.S. Patent No. #11,406,053)

2021 **Hu, J** & Ladoni, M., [Location Selection for Treatment Sampling](#) ---A field Study Design Tool to Optimize Treatment Assignment and Soil Sampling Locations for Model Calibration. (U.S. Patent No. #10,963,606)

2021 **Hu, J.**, [Location Selection for Model Assessment](#). (U.S. Patent No. #10, 990,716)

PUBLICATIONS

Public Peer-Reviewed Journal and Conference Articles

2023 **Hu, J.K.**, Tchetgen Tchetgen, E.J. and Dominici, F. [Using Negative Controls to Adjust for Unmeasured Confounding Bias in Time Series Studies](#). Nature Review Methods Primers 3, 66

2023 **Hu, J.K.**, Tchetgen Tchetgen, E.J. [Causal Inference with Time Series Data and Unmeasured Confounding](#). 39th Conference on Uncertainty in Artificial Intelligence

2023 **Hu, J. K.**, Zorzetto, D., & Dominici, F. [A Bayesian Nonparametric Method to Adjust for Unmeasured Confounding with Negative Controls](#). arXiv preprint arXiv:2309.02631. (Under Review for Bayesian Analysis)

- 2023 Aggarwal, S., **Hu, J. K.**, Sullivan, J. A., Parks, R. M., & Nethery, R. C. [Severe flooding and cause-specific hospitalization in the United States](#). arXiv preprint arXiv:2309.13142. (Under Review for Nature Communications)
- 2023 Spangler, K. R., Adams, Q. H., **Hu, J. K.**, Braun, D., Weinberger, K. R., Dominici, F., & Wellenius, G. A. [Does Choice of Outdoor Heat Metric Affect Heat-Related Epidemiologic Analyses in the US Medicare Population?](#). Environmental Epidemiology 7(4).
- 2023 Mauricio Tec, Ana Trisovic, Michelle Audirac, Sophie Mirabai Woodward, **Jie Kate Hu**, Naeem Khoshnevis, Francesca Dominici. “[SpaCE: The Spatial Confounding Environment](#)” arXiv preprint: arXiv 2312.00710 (Under Review for 41th International Conference on Machine Learning)
- 2021 **Jie Hu**, Norman E. Breslow, Gary Chan, Couper David. [Estimating the Hazard Difference from Case-Cohort Studies](#)”, European Journal of Epidemiology 36(11), 1129-1142.
- 2018 Norman Breslow and **Jie Hu**. [Survival Analysis of Case-Control Data: A Sample Survey Approach](#). Handbook of Statistical Methods for Case-Control Studies, Chapman and Hall/CRC.
- 2015 Norman E. Breslow, **Jie Hu**, Jon A. Wellner. [Z-estimation and Stratified Samples: Application to Survival Models](#). Lifetime Data Analysis 21, 493-516.
- 2014 **Jie Hu**, Xianlong Wang, Pei Wang. [Testing Gene-gene Interactions in Genome Wide Association Studies](#). Genetic Epidemiology 38, 123-134.
- 2016 Afzali, A., Park, C. J., Zhu, K., **Hu, J. K.**, Sharma, P., Sinanan, M. N., & Lee, S. D. [Preoperative Use of Methotrexate and the Risk of Early Postoperative Complications in Patients with Inflammatory Bowel Disease](#). Inflammatory Bowel Diseases 22(8), 1887-95.
- 2014 Afzali, A., Wheat, C. L., **Hu, J. K.**, Olerud, J. E., & Lee, S. D. [The Association of Poriasisiform Rash with anti-Tumor Necrosis Factor \(anti-TNF\) Therapy in Inflammatory Bowel Disease: A Single Academic Center Case Series](#). Journal of Crohn’s and Colitis 8(6), 480-488.
- 2013 Quade, S. J., Mourot, J., Afzali, A., Sinanan, M. N., Lee, S. D., **Hu, J. K.**, & Park, C. J. Mo1760 Assessment of Postoperative Complications in Patients With IBD. A Single Academic Medical Center Experience. Gastroenterology, 5(144), S-1109.

Under Review

Zixu Zhao, Melissa Lunden, **Jie Kate Hu**, Brian Lafranchi, Yutong Liang, Caleb Arata, Erin Katz, Allen H. Goldstein, & Haofei Zhang. “Air Pollution Mapping and Machine Learning Reveal Key Factors for Distinct Community-level Exposures in San Francisco, California” (submitted to *Environmental Science & Technology*)

In Preparation

- Jie Kate Hu**, “A Z-estimation System: A Modular Approach to Model Development.”
- Jie Kate Hu** & Gary Chan, “Weights Determination in Generalized Case-Cohort Study.”
- Jie Kate Hu**, & Eric Tchetgen Tchetgen, E.J, “Adjusting for Mismeasured Weather Time Series with Negative Controls for Effect Estimates.”
- Jie Kate Hu**, Ana Trisovic, Francesca Dominici, “Population Co-exposure to Extreme Heat, Wildfire, and Wildfire Smoke in Western US During the Past Decade.”
- Jie Kate Hu** & Ying Xu, “On-Farm Trial Design and Analysis to Estimate Field-Specific Treatment Effect with High-resolution Yield Monitor Data.”
- Julian Schmitt, Josh Yamamoto, **Jie Kate Hu**, Grayson White, George Gaines, Kelly McConville, “Zero Inflation in Small Area Estimation Models: Improving Forest Inventory Estimates.”

Peer-Reviewed Technical Reports, The Climate Corporation, Bayer Crop Science

- 2018 **Kate Hu**, Camila Casquilho & Megan Chen, “Combining Measurements and Models for Nitrogen Management.”
- 2018 **Kate Hu**, “Adjust Measurement-based Nitrogen Management Decisions using Biogeochemical Process Models.”
- 2018 Camila Casquilho, **Kate Hu** & Megan Chen, “A Bayesian Hierarchical Model for Critical Nitrate Estimation.”
- 2018 Carlos Carrion, **Kate Hu**, Andrew McGowan, Megan Chen, “Counterfactual Estimation of Yield Response as a Function of Soil Nutrients.”
- 2018 Camila Casquilho, **Kate Hu**, “Spatial Variability of Pre-sidedress Nitrate.”
- 2017 Zeshi Zheng, **Kate Hu**, Mike Malone, Nicholas Vogel, “A Time-Series Clustering Approach for Soil Moisture Probes Placement.”
- 2017 **Kate Hu**, “A Model-Assisted Probability Sampling Design for Representative Locations.”
- 2016 **Kate Hu**, “A Sampling Design for Model Assessment.”
- 2016 **Kate Hu**, Moslem Ladoni, “A Sampling and Treatment Placement Tool for the Climate Corporations Nitrogen Trials.”
- 2016 Jing Cao, **Kate Hu**, “Agronomist Survey Design and Analysis.”
- 2015 **Kate Hu**, “Evaluation of the 2015 Climatology Field Experiments.”
- 2015 **Kate Hu**, “Variable Seeding Rate Corn Research Partner Trials Analysis.”

PUBLIC SOFTWARE

- 2023 Zorzetto, D. & **Hu, J.K.** A Bayesian Nonparametric Method to Adjust for Unmeasured Confounding with Negative Controls. [Github Repo](#)
- 2022 VanLandschoot, M.J.D. , McConville J. Schmitte, K. & **Hu, J.K.** Informing Forestry professionals, Conservationists, and Regulators of the Potential Impacts of Climate Change on US Forests https://mjdvl.shinyapps.io/NCASI_APP/
- 2019 Ding, V. and **Hu, J.K.**, R Shiny app: [Additive Hazards](#)
- 2017 **Hu, J.K.**, Fit Additive Hazards Models for Survival Analysis, CRAN - Package [addhazard](#)

TEACHING EXPERIENCE

Harvard University, Mentor

Undergraduate Forestry Data Science Summer Projects (summer 2022, summer 2023)

Aclima, Inc. & University of Washington, Co-instructor

Biostatistics Data Science Capstone II (winter 2021)

Biostatistics Data Science Capstone I (fall 2020)

University of Washington, Teaching Assistant

Medical Biometry II (spring, 2012)

Harvard School of Public Health, Teaching Assistant

Principles of Biostatistics (fall, 2007)

STUDENT SUPERVISION

Doctoral Students

- 2022–present Dafne Zorzetto, Project: “Bayesian nonparametric methods to adjust for unmeasured confounding,” Statistics, University of Padova, Italy
- 2022–present Sarika Aggarwal, Project: “A national study of flood and its health impacts,” Biostatistics, Harvard University.
- 2017-2018 Zeshi Zheng, Project: “Machine learning for adaptive soil moisture sensor placement,” Civil Engineering, UC, Berkeley. Current data science manager at C3 AI

Master Student

- 2020-2021 Cathy Chen, Project “Developing health score based on hyperlocal air quality maps,” Biostatistics, University of Washington, Seattle. Currently data scientist at Abbot

Undergrad Student

- 2022 Maxwell J.D. VanLandschoot, Project “Visualizing climate change’s impact on US forests,” Economics, Reed College. Currently Schwarzman Scholar at Tsinghua University, summer
- 2022 Yibin Xiong, Project “Simulation studies of kernel methods for unmeasured confounding bias adjustment,” Applied Mathematics, USC. Currently MS student in Biostatistics at Harvard
- 2022 Josh Yamamoto, Project “Zero Inflation in Small Area Estimation Models: Improving Forest Inventory Estimates,” Statistics and Mathematics, Reed College. Currently Statistical Software Engineer at U.S. Forest Service, summer
- 2022 Julian Schmitte, Project “Zero Inflation in Small Area Estimation Models: Improving Forest Inventory Estimates,” Applied Mathematics, Harvard College. Currently PhD student in Environmental Science and Engineering at Caltech, summer

INVITED TALKS

- 2023 Precision and Digital Agriculture, Guest Speaker in MBA course Agribusiness, Harvard Business School, Boston, MA, Sep
- 2023 Leverage Proxies to Adjust for Unmeasured Confounding, Alamo Statistics Symposium, San Antonio, TX, Apr
- 2022 Discussant on “Disentangling Confounding and Dependence in Spatial Statistics”, Causal Inference Seminars, Harvard Data Science Initiative, Boston, MA, Oct
- 2022 Use Negative Controls to Adjust for Unmeasured Confounding with Time Series Studies in Environmental Epidemiology, 35th New England Statistics Symposium, Storrs, CT, May
- 2021 Air Quality Assessment from Mobile Sensing Platform, Electrical and Computer Engineering, Stony Brook University, May
- 2017 Hypothesis Formulation, Experimental Design, and Analysis of Precision Agriculture Trials, Women in Statistics and Data Science, La Jolla, CA, October
- 2017 Stories of Success, Lessons Learned, and Advice for Productive and Enjoyable Collaborations, Chair, Panel Talk, Women in Statistics and Data Science, La Jolla, CA, October
- 2016 Using the Additive Hazards Model with Two-Phase Sampling in Atherosclerosis Risk in Community Study, John Hopkins University, September

2014 Z-estimation for General Two-phase Sampling Problems, National Cancer Institute, Bethesda, MD, June

CONFERENCE TALKS

- 2023 Causal Inference with Time Series Data and Unmeasured Confounding, 39th Conference on Uncertainty in Artificial Intelligence, Pittsburg, PA, July
- 2023 A Bayesian Nonparametric Method to Adjust for Unmeasured Confounding with Negative Controls, Joint Statistical Meetings, Toronto, Canada, August
- 2020 Representative Sampling Method for Air Quality Monitoring with Mobile Sensing Platform, 113th Air & Waste Management Association, San Francisco, CA, June
- 2018 Application of Causal Bayesian Networks to Environmental Data, Atlantic Causal Inference Conference, Pittsburgh, PA, May
- 2013 Application of Z-estimation Theory to Calibrated Estimators for Semi-parametric Models with Two-phase Stratified Sampling, Graybill Conference on Modern Survey Statistics, Fort Collins, CO, June
- 2012 Parametric and Semi-parametric Analysis of Mean Residual Life Acceleration, 8th World Congress on Probability and Statistics, Istanbul, Turkey, June
- 2011 Comparison of Tests for Association with Interaction of Genes Between Two Loci, Joint Statistical Meetings, Miami, Florida, August

PROFESSIONAL SERVICE

Reviewers

NeurIPS, Computational Statistics and Data Analysis, Lifetime Data Analysis, BMC Genetics Editorial Board, *Journal of Environmental Quality*

To University

Student-Faculty Communications Committee, University of Washington, 2012-2013
Computer-Policy Committee, University of Washington, 2008-2009

To Community

Co-chair, University of Washington Statistics in the Community, 2012-2014

COMPUTING LANGUAGES

Python, R, Julia, Spark, SAS, STATA, UNIX, Scala, C++, Mathematica, Git, LaTeX, Markdown, Sphinx, PlantUML, BigQuery SQL, Google Cloud, Google Earth, AWS, Docker, Jupyter

REFERENCES

Eric Tchetgen Tchetgen
Luddy Family President's Distinguished Professor, Statistics and Data Science
University of Pennsylvania
Email: ett@wharton.upenn.edu

Francesca Dominici
Clarence James Gamble Professor of Biostatistics, Population and Data Science
Harvard T.H. Chan School of Public Health
Director of the Harvard Data Science Initiative
Email: fdominic@hsph.harvard.edu

Gary Chan
Professor, Biostatistics
University of Washington
Email: kcgchan@uw.edu

Steve Sain
Senior Director and ASA fellow, Geospatial and Data Science
Jupiter Intelligence
E-mail: sainsr2@gmail.com

Adam. A. Szpiro (Teaching)
Professor, Biostatistics
University of Washington
Email. aszpiro@uw.edu