Kate Jie Hu

206-910-3140 | contact@katehu.com | linkedin.com/in/jie-kate-hu-1342157/

Current Position

Harvard University 11/2021 - present

Postdoctoral Fellow, Advisor: Francesca Dominici

Education

University of Washington

PhD Biostatistics *2014*

- Thesis: "A Z-estimation System for Two-phase Sampling with Applications to Additive Hazards Models and Epidemiologic Studies"
- Advisors: Norman Breslow and Gary Chan

Harvard University

MS Biostatistics *2008*

University of Hong Kong

BSc Biochemistry, First Class Honors *2006*

Teaching

Medical Biometry II

Teaching Assistant @ U. of Washington *2012*

Principles of Biostatistics

Teaching Assistant @ Harvard School of Public Health *2007*

Professional Experience

Head of Data Science/Principal Data Scientist

Aclima Inc, San Francisco *2/2019 - 5/2021*

Use the mobile sensing platform and cloud-based analytics platform to turn big data about air quality into hyperlocal insights

- Built a multidisciplinary data science team from ground zero in 2019 including 5 PhD and 2 MS members; ranked No.1 in the 10 most innovative data science team by Fast Company in 2021
- Co-designed and sponsored a capstone course with instructors from the University of Washington: "using hyper-local air pollution data to advance environmental justice"; mentored Postdoctoral, PhD, and MS-level data scientists, students, and interns
- Developed and executed long-term research road maps for the data science team. The research program included big data collection and analytics, sampling and experimental designs, machine learning, uncertainty quantification, measurement errors, operation research, and spatiotemporal models
- Led R&D on sampling designs for mapping block-by-block air quality with mobile sensing platforms, deployed in multiple metropolitan areas
- Led R&D on inferential statistical tools, assisted by machine learning, used by several regulatory agencies and environmental justice communities
- Forged collaborations with atmospheric scientists, data scientists, and medical researchers across academia, governments, and industry

Senior Quantitative Researcher

The Climate Corporation, San Francisco *12/2014 - 2/2019*

Research program lead in using dynamic treatment regimes for precision agriculture *08/2017 - 2/2019*

- Secured budget for several FTEs and large-scale field trials over five states

- Led data scientists, data engineers, and environmental scientists to develop and test precision treatment decision models. Research areas included developing location-specific treatment algorithms based on the combinations of in-season climate and soil measurements, biogeochemical process models, machine learning, and causal inference; measurement variability studies and measurement protocol development; designing and conducting agricultural efficacy trials on both research and non-research farms for precision treatment algorithms evaluation

Research program lead in experimental and sampling design *01/2016 - 08/2017*

- Research areas included estimating average treatment effects with observational data; adapting existing field trials data to new hypotheses testing; adaptive sampling designs for in-ground weather and soil sensors placement; sampling designs for model calibration and validation; developed an internal experiment design R package TCCDesign, being used by environmental scientists in two consecutive years to collect field data for model improvement
- Lead authors/co-authors of 10+ internal peer-reviewed research papers and three patents

Research Assistant

Fred Hutchinson Cancer Research Center, Seattle *2008 - 2014*

- Developed new semi-parametric hazards models with applications to a HIV/AIDS study
- Studied definitions and tests of gene-gene interaction for Genome Wide Association Studies
- Evaluated the prediction capability of risk predictors and examined biases of ROC curve

Consultant

School of Medicine, University of Washington, Seattle *2012 - 2014*

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

Book Chapter

- Norman Breslow and Jie Kate Hu, "Survival Analysis of Case-Control Data: A Sample Survey Approach", *Handbook of Statistical Methods for Case-Control Studies*, Chapman and Hall/CRC, (2018)

Peer-Reviewed Journal Publications

- Jie Hu, Norman E. Breslow, Chan Gary, Couper David, "Estimating Disease Hazard Differences from Case-Cohort Studies", *European Journal of Epidemiology* **Jun**, 1-14 (2021).
- Afzali Anita, Christopher J. Park, Kehao Zhu, Jie Kate Hu, Prachi Sharma, Mika N. Sinanan, and Scott D. Lee, "Preoperative Use of Methotrexate and the Risk of Early Postoperative Complications in Patients with Inflammatory Bowel Disease", *Inflammatory Bowel Diseases* **22(8)**, 1887-95 (2016).
- Norman E. Breslow, Jie Hu, Jon A. Wellner, "Z-estimation and Stratified Samples: Application to Survival Models", *Lifetime Data Analysis* **21**, 493-516 (2015).
- Jie K. Hu, Xianlong Wang, Pei Wang, "Testing Gene-gene Interactions in Genome Wide Association Studies", *Genetic Epidemiology* **38**, 123-134 (2014).
- Afzali Anita, Chelle L. Wheat, Jie K. Hu, John E. Olerud, and Scott D. Lee, "The Association of Psoriasiform 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 (2014).
- Manickavasagan,Hanisha, Butnariu Madalina, Porter Kyle , Hu K. Jie, Husain Syed, and Afzali Anita"Inflammatory Bowel Disease Patients with Type 2 Diabetes and Obesity have a Higher Annual Burden and Costs of Hospitalizations: A call for action"(under review)
- Zhao Zixu, Lunden Melissa, LaFranchi Brian; Hu Jie; Zhang HaofeiHigh-Resolution Investigation of Spatiotemporal Variations of Urban Air Pollutants in San Francisco, California Using Mobile Monitoring and Big Data Analytics(under review)

Selected The Climate Corporation Internal Peer-Reviewed Technical Reports

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

Patents

- Jie Hu, Sampling Method for Mobile Monitoring Platforms. (Pending)
- Jie Hu, Carlos Carrion, Using Causal Learning Algorithms to Assist in Agricultural Management Decisions. (Pending)
- Jie Hu, Moslem Ladoni, Location Selection for Treatment Sampling. (US20200184128A1)
- Jie Hu, Location Selection for Model Assessment. (US20180260504A1)

Public R Package

- Jie Hu, Fit Additive Hazards Models for Survival Analysis. CRAN - Package addhazard

Selected Talks

- Representative Sampling Method for Air Quality Monitoring with Mobile Sensing Platform, Air & Waste Management Association 113th Annual Conference, San Francisco, CA, 06/20
- Application of Causal Bayesian Networks to Environmental Data, Poster, Atlantic Causal Inference Conference, Pittsburgh, PA, 05/18
- Hypothesis Formulation, Experimental Design, and Analysis of the Precision Agriculture Trial, Invited Talk, Women in Statistics and Data Science, La Jolla, CA, 10/17
- Stories of Success, Lessons Learned, and Advice for Productive and Enjoyable Collaborations, Chair of the Panel Talk, Women in Statistics and Data Science, La Jolla, CA, 10/17

- Using the Additive Hazards Model with Two-Phase Sampling in Atherosclerosis Risk in Community Study, Invited Talk, John Hopkins University, Baltimore, MD, 9/16
- 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, 6/13
- Parametric and Semi-parametric Analysis of Mean Residual Life Acceleration, 8th World Congress on Probability and Statistics, Istanbul, Turkey, 6/12

Computing Skills

- Experienced with R, Python, Git, Spark, SAS, LaTeX, STATA, UNIX, Scala
- Familiar with C++, Mathematica

Service

- Co-chair: University of Washington Statistics in the Community (2012-2014)
- Member: Student-Faculty Communications Committee, Department of Biostatistics at the University of Washington (2012-2013)
- Member: Computer-Policy Committee, Department of Biostatistics at the University of Washington (2008-2009)

Recreational Activities

- Halberstadt Fencers' Club, San Francisco
- The Mountaineers Club, Seattle
- Intramural Champions: Badminton Womens Doubles (2010, 2012), Co-Rec Softball (2012)