

Xiaotian (Kevin) Wu

CONTACT INFORMATION	210 Waterman St. Apt 11, Providence RI, 02906 Tel: 203-988-2098 Email: xiaotian_wu@brown.edu Personal Site: wuxiaotiankevin.github.io
SUMMARY	Passionate and skilled statistician with 4+ years of collaboration experience with local hospitals. Primarily specialized in developing novel statistical methods for single cell RNA sequencing data, with broad training in clinical trials. Solid knowledge and hands on experience in statistical genomics, convex optimization, topic modeling, state space modeling and G computation gained along PhD training and multiple collaborations.
EDUCATION	Brown University , Providence, RI Ph.D. Candidate, Biostatistics, <i>Expected</i> : Winter 2019 <ul style="list-style-type: none">• Advisor: Zhijin Wu, Ph.D• Thesis Topic: <i>Statistical Analysis of Single Cell RNA Sequencing Data</i> Yale University , New Haven, CT M.S., Biostatistics, May 2015 <ul style="list-style-type: none">• Advisors: Michael Wineringer, Ph.D and Peter Peduzzi, Ph.D• Thesis Topic: <i>Effectiveness of Intensive Therapy in Restoration of Motor Function in Chronic Stroke</i>• Equivalent GPA: 3.87 / 4.0 Hong Kong University of Science and Technology , Hong Kong SAR B.S., Biology, May 2013 <ul style="list-style-type: none">• Advisor: Professor W.K.R. Wong• Thesis Topic: <i>Extracellular Expression of Heterologous Proteins in Escherichia coli</i>
SKILLS	Programming <ul style="list-style-type: none">• R, Python, SAS, SQL, C++, MATLAB, Stata, L^AT_EX, Bash Statistics <ul style="list-style-type: none">• Statistical Inference, Convex Optimization, Machine Learning, Topic Models, Causal Inference, Algorithms and Data Structures, and others
SELECTED RESEARCH EXPERIENCE	Research Assistant 2015 - Present Department of Biostatistics, Brown University Project 1: Penalized Latent Dirichlet Allocation for Single Cell RNA Sequencing Data **ENAR award-winning project** Supervisor: Zhijin (Jean) Wu, Ph.D <ul style="list-style-type: none">• Develop a Latent Dirichlet Allocation (LDA) model with biological interpretable penalization for analyzing single cell RNA sequencing data.• Prove convexity of the problem. Derive and implement convex optimization algorithm for penalized LDA parameters estimation. Build R package utilizing Rcpp to boost speed.• Apply the method to simulated and real single cell RNA sequencing data. The method demonstrates superior classification accuracy and provides interpretable dimensionality reduction results. Project 2: A State Transition Framework for Patient-level Modeling of Engagement and Retention in HIV Care Using Longitudinal Cohort Data Supervisor: Hana Lee, Ph.D and Joseph Hogan, Ph.D <ul style="list-style-type: none">• Explore longitudinal cohort HIV data. Discuss and decide on proper HIV state definition with collaborators.• Combine data from multiple source. Preprocess data and encode proper state for patients.• Model data with the proposed state transition framework. Implement a parallel version of G computation to predict treatment effect under different treatment regimen.

Project 3: **Mouse Ovary Single Cell RNA Sequencing Analysis**

Supervisors: Lynae Brayboy, MD and Zhijin (Jean) Wu, Ph.D

- Perform single cell RNA sequencing exploratory data analysis for quality control.
- Identify differentially expressed genes. Perform gene ontology analysis. Extract gene information using web scrapping.

Research Assistant

2014 - 2015

Department of Biostatistics, Yale University

Cooperative Studies Program Coordinating Center, Department of Veterans Affairs at West Haven

Project 4: **Effectiveness of Intensive Therapy in Restoration of Motor Function in Chronic Stroke**

Supervisors: Michael Winger, Ph.D and Peter Peduzzi, Ph.D

- Perform subgroup analysis on robotics rehabilitation clinical trial data. Identify subgroups who benefit more from treatment.
- Cluster kinematics data from robotics rehabilitation clinical trial with a greedy method.
- Predict patient outcome from kinematics data obtained by robot rehabilitation equipment.

CONSULTING EXPERIENCE

Transfer Rate and Live Birth Rate Prediction for *In Vitro* Fertilization

Collaborator: Shunping Wang, Ph.D, HCLD, Women & Infants Hospital of Rhode Island

- Introduce and explain state of the art machine learning prediction model to clinical audience.
- Implement xgboost for *in vitro* fertilization (IVF) transfer and live birth prediction.
- Build and host R Shiny web application for IVF data visualization and online, real time prediction.
- Connect doctors with a local research facility for analyzing embryo imaging data with auto-encoder for feature extraction.

HONORS & AWARDS

Scholarship — American Federation for Aging Research (AFAR)

- AFAR Scholarship for Research in the Biology of Aging 2018

Presentation Award — Department of Biostatistics, Brown University

- PhD Research Presentation Award 2018

Student Paper Award — Eastern North American Region (ENAR)

- ENAR Distinguished Student Paper Award 2018

Scholarship — Department of Biostatistics, Yale University

- The Dr. Colin White Memorial Scholarship 2014

Scholarship — Hong Kong University of Science and Technology

- The Joseph Lau Luen Hung Charitable Trust Scholarship 2009

PUBLICATIONS

In Preparation

1. **Xiaotian Wu**, Hao Wu, Zhijin Wu. “Penalized Latent Dirichlet Allocation Model in Single Cell RNA Sequencing” ****ENAR award-winning paper****

Published

1. Genberg, Becky L, Hana Lee, Joseph W Hogan, Fatma Some, Juddy Wachira, **Xiaotian K Wu**, and Paula Braitstein. 2018. “Point of Diagnosis and Patient Retention in Hiv Care in Western Kenya.” *JAIDS Journal of Acquired Immune Deficiency Syndromes* 78 (4). LWW: 383–89.
2. Lee, Hana, Joseph W Hogan, Becky L Genberg, **Xiaotian K Wu**, Beverly S Musick, Ann Mwangi, and Paula Braitstein. 2018. “A State Transition Framework for Patient-Level Modeling of Engagement and Retention in Hiv Care Using Longitudinal Cohort Data.” *Statistics in Medicine* 37 (2): 302–19.
3. Lee, Hana, **Xiaotian K Wu**, Becky L Genberg, Michael J Mugavero, Stephen R Cole, Bryan Lau, and Joseph W Hogan. 2018. “Beyond Binary Retention in Hiv Care: Predictors of the Dynamic Processes of Patient Engagement, Disengagement, and Re-Entry into Care in a Us Clinical Cohort.” *AIDS (London, England)*.

4. Li, Ling, John Hartigan, Peter Peduzzi, Peter Guarino, Alexander Beed, **Xiaotian Wu**, and Michael Wining. 2018. "Clustering of Directions Improves Goodness of Fit in Kinematic Data Collected in the Transverse Plane During Robot-Assisted Rehabilitation of Stroke Patients." *Frontiers in Robotics and AI* 5. Frontiers: 57.
5. Papas, Rebecca K, Benson N Gakinya, Michael M Mwaniki, **Xiaotian K Wu**, Hana Lee, Steve Martino, Debra A Klein, et al. 2018. "Associations with Unprotected Sexual Behavior Among Hiv-Infected Drinkers in Western Kenya." *AIDS and Behavior*. Springer, 1–11.
6. **Wu, Xiaotian**, Peter Guarino, Albert C Lo, Peter Peduzzi, and Michael Wining. 2016. "Long-Term Effectiveness of Intensive Therapy in Chronic Stroke." *Neurorehabilitation and Neural Repair* 30 (6). SAGE Publications Sage CA: Los Angeles, CA: 583–90.

PRESENTATIONS	<p>Poster, "Penalized Latent Dirichlet Allocation Model in Single Cell RNA Sequencing", Joint Statistical Meeting, Vancouver, British Columbia, Canada, July 2018</p> <p>Contributed, "Penalized Latent Dirichlet Allocation Model in Single Cell RNA Sequencing", ENAR Spring Meeting, Atlanta, GA, March 2018</p> <p>Poster, "A Robust Estimator for Poisson Mean with Zero-inflation and Contamination", Joint Statistical Meeting, Baltimore, MA, August 2017</p> <p>Contributed, "State Space Models of Retention, Disengagement, and Re-entry into HIV Care", ENAR Sprint Meeting, Washington D.C., March 2017</p>	
TEACHING EXPERIENCE	<p>Teaching Assistant Fall 2017–18 PHP2250 - Advanced Quantitative Methods in Epidemiologic Research Instructor: Chanelle J Howe, Ph.D Department of Epidemiology, Brown University</p> <ul style="list-style-type: none"> • Lectured linear and generalized linear models for confounding adjustment. • Lectured Cox regression model for confounding adjustment. <p>Teaching Assistant Fall 2016–17 PHP2550 - Practical Data Analysis Instructor: Christopher Schmid, Ph.D Department of Biostatistics, Brown University</p> <ul style="list-style-type: none"> • Lectured classification methods. • Led lab sessions on practical data analysis using R. <p>Teaching Assistant Fall 2015–16 PHP2520 - Statistical Inference I Instructor: Zhijin Wu, Ph.D Department of Biostatistics, Brown University</p>	
MEMBERSHIP IN PROFESSIONAL SOCIETIES	<p>American Statistical Association</p> <p>International Chinese Statistics Association</p>	
SERVICE	<p>Member, Diversity and Inclusion Committee Sept 2017 – May 2018 Department of Biostatistics, Brown University</p> <p>Student Representative Sept 2016 – Aug 2017 Department of Biostatistics, Brown University</p> <p>President, Biostatistics Students Association May 2014 – May 2015 Department of Biostatistics, Yale University</p>	
INDUSTRY EXPERIENCE	<p>Yelp Inc, Quantitative Researcher/Statistician Summer Intern Jun 2018 – Aug 2018</p> <ul style="list-style-type: none"> • Combine data from multiple sources using Redshift. • Prototype and productionize an upsell account recommendation model to identify upsell opportunities for sales representatives using Python. • Explore personalized rating for restaurant recommendation using latent features from collaborative filtering. 	

FOR FUN

Kaggle Competition

TalkingData Mobile User Demographics, 88th of 1689, top 6%

Jul 2016 – Sept 2016

- Merge multiple large datasets and perform feature engineering.
- Apply xgboost and neural network for multi-group classification.