Xiaoxuan Cai

xiaoxuan.cai@yale.edu | Tel. 203-843-0518
https://xiaoxuan-cai.github.io

EDUCATION

Yale University, Department of Biostatistics **New Haven, CT**
Ph.D. in Biostatistics, MS in Biostatistics *Expected May 2020*

- **Concentrations:** Causal Inference, Survival analysis
- **Courses:** Stochastic Processes, Models of Operations Research and Management, Bayesian Statistics, Computational Statistics, Nonparametric Statistics, Statistical Inference, Generalized Linear Model

Peking University, School of Life Science **Beijing, China**
Bachelor of Sciences in Biological Science, Bachelor of Arts in Economics *July 2013*

- **Courses:** Genetics, Biochemistry, Molecular Biology, Cell Biology, Physiology, Functions of Real Variables, Econometrics, Algorithm and Data Structures and Computer Operation, Economics of Information, Health Economics, Environmental Economics, Microeconomics, Macroeconomics

ANALYTICAL EXPERIENCE

Department of Biostatistics (Advisor: Forrest W. Crawford), Yale University **New Haven, CT**
PhD Research *Dec 2015 – Present*

Causal identification for the infectious disease intervention under contagion

- Developed an innovative two-stage causal identification strategy for infectious disease outcomes under contagion
- Developed novel non-parametric methods for identification, using the theory of competing risks. The approach makes no requirements on study design or the model of transmission probability, making it widely applicable to infectious disease data.
- Used the new framework to examine the causal validity of popular estimands in epidemiology

Measuring the effects of infectious disease interventions in networks

- Proposed novel estimators of direct and indirect vaccine effects based on transmission hazard, by extending the causal inference framework to clusters and integrating dynamic modeling of disease transmission.
- Developed parametric and semi-parametric models for parameter estimation under a particular network structure, using counting process. Proved the consistency of the estimands under martingale central limit theory.
- Generated simulations in clusters of size two up to ten under various dynamics of disease transmission, and examined the performance of our estimands and other popular estimands in epidemiology

Takeda Pharmaceutical Company **Cambridge, MA**
Summer Intern to the Statistics Department *June 2016 – Aug 2016*

- Implemented methods for mining Twitter data (~300,000 tweets) to identify the most influential users for a given topic
- Implemented a topic classification algorithm for extracting main topics from a collection of raw Twitter messages; Adapted the method to use the Twitter API to perform real-time analysis
- Led discussions on text mining and information extraction techniques in weekly department meetings
- Presented summer project to the Takeda Statistics Department with an audience of over 100 people

Department of Political Science (Dr. Peter Aronow Lab), Yale University **New Haven, CT**
Graduate Student Researcher *Oct 2014 – May 2015*

- Applied LASSO variable selection method to Generalized Estimation Equation (GEE) model for binary response variable, and write R codes for implementation.
- Applied to longitudinal dataset about cocaine addiction during pregnancy, and compare analysis results with the GEE model for negative binomial distribution both in real data and simulations

School of Medicine, (Dr. Kimberly Yonkers Lab), Yale University **New Haven, CT**
Graduate Student Researcher *Mar 2014 – Apr 2015*

- Utilized multiple analytical models to identify risk factors for cocaine addiction among pregnant woman:
  - Mixed effects regression model for ordinal outcomes
  - Generalized estimation equation (GEE) model for ordinal outcomes and GEE model for negative binomial distribution
  - Zero-inflated and zero-altered Poisson model for correlated data
- For each model, performed single variable exploration and backward selection.
- Identified age of first use, social support, baseline cocaine use, marijuana usage as key potential risk factors

Yale Center of Analytical Science (Dr. Peter Peduzzi Lab), Yale University **New Haven, CT**
Graduate Student Researcher *May 2014 – Aug 2014*

- Generated data under different prior distributions and different link functions, and used the posterior mean or other suitable estimators to decide the proper dose level of phase II trial.
- Compared the patients' assigned dosages with those determined by the traditional 3+3 method.

PUBLICATIONS

Xiaoxuan Cai, Wen Wei Loh, and Forrest W. Crawford (2019) Identification of causal intervention effects under contagion – *Submitted* (https://arxiv.org/abs/1912.04151)
Regina Melendez, **Xiaoxuan Cai,** Cristine Hine, et al. (2015) Correlates of Cocaine Use in Pregnancy. Yale Medicine Thesis Digital Library.

PRESENTATIONS

2019 Research in progress presentation, School of Public Health, Yale University, New Haven, CT
2019 Contributed presentation, 2019 Joint Statistical Meetings, Denver, CO
2019 Invited presentation, 33nd New England Statistics Symposium, Hartford, CT
2019 Invited session, 2019 ENAR, Philadelphia, PA
2018 Invited presentation, 2018 Women in Statistics and Data Science Conference, Cincinnati, OH
2018 Poster session, 32nd New England Statistics Symposium, Amherst, MA
2018 Speed presentation, 2018 Joint Statistical Meetings, Vancouver, BC

AWARDS AND RECOGNITION

2018 Travel Award, Women in Statistics and Data Science conference
2018 Travel Award, Summer Institutes at the University of Washington
2018 MassMutual Student Poster Award, 32nd New England Statistics Symposium
2015 Fellowship from Takeda pharmaceutical Company
2013 Huirong Li Scholarship, Peking University

LEADERSHIP and TEACHING EXPERIENCE

LEADERSHIP

Chairperson of the Connecticut Peking University Alumni, May 2017 to Sep 2018
Class President, Department of Life Sciences, Peking University, Sep 2010 to Jun 2012
Minister of the Academic Department, the Student Union, Peking University, Sep 2010 to Jul 2011
Member of the School's Debate Team, Peking University, Sep 2009 to Jul 2010

TEACHING EXPERIENCE

Yale University, Department of Biostatistics **New Haven, CT**
Teaching Assistant (Course: Applied Survival Analysis, Introduction to Biostatistics II, Applied Regression Analysis, Biostatistics in Clinical Investigation) *Aug 2014 – May 2018*

- Led weekly office hours and lab sessions; graded midterms and finals

TRAININGS AND SKILLS

Trainings: Causal Inference and Big Data Summer Camp, University of Pennsylvania (2017)
Summer Institute in Statistics and Modeling in infectious disease, University of Washington (2018)
Skills: Python, R, C, C++, SPSS, Stata, Matlab, Git, LaTeX;
Language: Chinese Mandarin (native), English (Fluent)