

# ZHUOHUI LIANG

WORK PHONE : +1 (646) 496-3283 | EMAIL 1 : zl2974@cumc.columbia.edu | EMAIL 2 : jeffreyltang1995@outlook.com

## EDUCATION

- **INSTITUTE : COLUMBIA UNIVERSITY** New York City, U.S.  
MAJOR : MASTER CANDIDATE IN BIOSTATISTICS(4.07/4.3) 09/2020-PRESENT
- **INSTITUTE : SHANDONG UNIVERSITY** Jinan, China  
MAJOR : BACHELOR OF PUBLIC HEALTH(4.32/5) 09/2014-06/2019

## RESEARCH EXPERIENCE

- **Department of Biostatistics, Columbia University** New York  
*Research Assistant, Advisor: Prof. Shuang Wang* 04/2021-PRESENT  
**Methodology Developing on Multi-omic Association and Clustering**
  - Literature review on dimensional reduction, discrete time random walk and multi-view clustering and association singular boosting methods
  - Design simulation to test and compare existing multi-view clustering method, ie. Similarity Network Fusion(SNF), Multi-view Clustering Without Parameter(COMIC), Partition-level Multiview Subspace Clustering(PSMC), Diversity-Induce Multiview Clustering(DiMSC) and etc.; and association boosting method, ie. Disease-Specific Network Enhancement Prioritization(DISNEP), Gene-wanderer etc.
  - Experiment above-mentioned methods with alternative manifold learning for new data representation, smoothing learned similarity matrix with t-step self random walk, multi-view fusion or consensus, complementary regulation on clustering result to develop a method that can identify true number of clusters and clustering result under scenario which different view provide information only to find partial subtypes.
- **Department of Biostatistics, Columbia University** New York  
*Research Assistant, Advisor: Prof. Ying Wei, Prof. Yifei Sun* 09/2021-PRESENT  
**Predicting Patient Adherence with Multivariate Random Forest**
  - Design and code weighted multivariate random forest algorithm independently to capture the within interval pattern and predict medication adherence for an incomplete medication adherence history data;
  - Model missing pattern of the response variables under Missing At Random(MAR) assumption considering the missingness was mostly due to researchers' decisions. And boost compliance prediction model's performance with the inverse predicted missing observation probability(IPW) and see variables importance of baseline hospitalization vital scores and mental status scores;
  - Ensemble all trees with Hidden Markov Chains model for a dynamic prediction model;
  - Design simulation study to test algorithm performance and analysis on the data, we found that using IPW to adjust for missing data improves the prediction performance of our multivariate random forest under MAR in simulation and our adherence history data;
  - Manuscript writing
- **Department of Social Health, Shandong University** Jinan, China  
*Research Assistant, Advisor: Prof. Jiajia Li* 01/2019-06/2019  
**Case Study: eHealth Resource Allocation Inequality in Shandong Province**
  - collect data on the use of electronic health (eHealth) resources from an online healthcare community involving more than 441,000 doctors and 31 provinces in China, by using Python packages, such as lxml and pandas.
  - Cleaned and visualize data with R-ggplot; utilized techniques such as visualizing the spatial distribution and density of highly-ranked hospitals and experienced professionals on map.

- Performed statistical analysis on eHealth online-to-offline spillover effects using the Spatial Durbin Panel Model to reduce autocorrelation between regions;
- Identified the spillover effects of eHealth services on health care inequity between regions; provided suggestions and strategies to address healthcare disparities using widely accessible eHealth resources.
- Drafted and submitted a report to the 2020 Health Technology Assessment International Conference and 2019 Global Young Scholars Forum on Health Management

- **Department of Social Health, Shandong University**

Jinan, China

*Research Assistant, Advisor: Prof. Jiajia Li*

10/2018-06/2019

**Project: Effect of Insurance Plans on Catastrophic Healthcare Expenditures**

- Led a research group to conduct a study on the effects of different insurance types on catastrophic healthcare expenditures.
- Conducted data cleansing using R package – tidyverse; processed information from 44,417 observational sessions and identified 3,300 missing units of data and outliers.
- Collected geographical locations based on provided address with Google Maps API and matched location with the insurance coverage provided by each plan as an instrumental variable for addressing the endogeneity introduced by insurance selection bias.
- Performed two-stage least square regression analysis to build a model with instrumental variables using R package -AER; concluded that China's new health insurance program has decreased the odds of catastrophic health expenditures but has failed to control total health expenditures.
- Provided quantitative evidence to forecast the financial expenditures of different healthcare insurance plans and to provide a reference for other researchers in the selection of instrumental variables.

## PUBLICATION

Li,J., Li,S., Cao,W., Wang,Z., **Liang,Z.**, Fu, W., and Zhao, J. (2021). Appraisal of china's response to the outbreak of covid-19 in comparison with sars. *Frontiers in Public Health*, 9:830

## WORK EXPERIENCE

- Mailman School of Public Health-Columbia University

New York

*Teaching Assistant* **High-Dimensional EHR data Bootcamp**

09/2021-PRESENT

- Course planning, material preparation and teaching lab session for MIMIC Data Preparation and Sparse Coding and Instrumental Variable Analysis Method for **High-Dimensional EHR data Bootcamp**

- Mailman School of Public Health-Columbia University

New York

*Teaching Assistant* **Data Science I P8105**

06/2021-09/2021

- Course planning, holding office hour and grading for **Data Science I**

- National Health Development and Research Center of China Health Committee

Beijing, China

*INTERN*

06/2019-06/2020

- Provide advice on sampling method, model selection, and result interpretation as well as data visualization.
- Estimate national health expenditure in 2019, and analyze the distribution of different demographic characteristics, such as age and gender, to optimize healthcare resource allocation.
- Build ICD-10 auto-coding and data format auto-standardizing algorithm with Python for annual health accounting, improving working efficiency, data accuracy, and integrity.
- Conduct a literature review on data mining methods for detecting healthcare fraud; provide data support for a pilot feasibility project.
- Participated in data visualization of healthcare metrics and healthcare access for *Development Report On Health Reform In China(2020)*(ISBN:9787520164412).
- Assisted with study design and data processing in Python for two confidential projects

## SKILLS

- Computing: Python, R, Linux command line