

# ZHUOHUI LIANG

More info: [jeffreyliang-vandy.github.io](https://jeffreyliang-vandy.github.io) | EMAIL : [zhuohui.liang@vanderbilt.edu](mailto:zhuohui.liang@vanderbilt.edu) | [Linkedin](#)

## EDUCATION

- **VANDERBILT UNIVERSITY** Nashville, U.S  
MAJOR : DOCTORAL PROGRAM IN BIOSTATISTICS 08/2022-PRESENT
- **COLUMBIA UNIVERSITY** New York City, U.S.  
MAJOR : MASTER OF SCIENCE IN BIOSTATISTICS(4.07/4.3) 09/2020-06/2022
- **SHANDONG UNIVERSITY** Jinan, China  
MAJOR : BACHELOR OF PUBLIC HEALTH(4.32/5) 09/2014-06/2019

## RESEARCH EXPERIENCE

- **Department of Biostatistics, Vanderbilt University** Nashville  
*Research Assistant, Advisor: Prof. Bryan E.Shepherd, Chao Yan, Prof. Marlin A.Bradley* 09/2023-PRESENT  
**HIV cohort synthesis data generation**
  - Goal:Generate publishable HIV-cohort EHR data with Generative Adversary Network.
  - Explore methods for synthetic data generation, including conditional model simulation, Variational Auto-encoder. And eventually collaborate with Prof. Bradley and Chao Yan to develop a Generative Adversary Network to generate HIV cohort data.
  - A synthetic cross-sectional baseline cohort data has been generated and validated with test data, we observed great similarity to real data and applicable to real study and reproduce the result similar to real data.
  - Longitudinal version of the data is under development, along with assisting R01 application.
- **Department of Biostatistics, Vanderbilt University** Nashville  
*Research Assistant, Advisor: Prof. Bryan E.Shepherd* 09/2022-06/2023  
**Trend and baseline factors associated with time to HAART treatment in HIV baseline TB patient**
  - Worked closely with bioinformatitian to generate study cohort from 2006-2020 from CCASANET(Caribbean,Central, and South America Network for HIV Epidemiology)
  - Data analysis for the 30k patient cohort to identify changes in trend of HAART starting time among HIV patient with TB at enrollment, compared to non-TB patient
  - Trend was analysis with cox proportional hazard model, we discover a general declining trend in all site, but Haiti has unreasonable trend after analysis
  - Working with collaborator, we identify potential data collection mistake in Haiti and working with sites to correct the EHR database.
- **Department of Biostatistics, Columbia University** New York  
*Research Assistant, Advisor: Prof. Shuang Wang* 04/2021-05/2022  
**Methodology Developing on Multi-omic Clustering**
  - Literature review on data integration methods and multi-omics clustering
  - Design simulation to test and compare existing multi-omics clustering method, ie. Similarity Network Fusion(SNF), and co-training based Multi-view Clustering Without Parameter(COMIC), Partition-level Multiview Subspace Clustering(PSMC), Diversity-Induce Multiview Clustering(DiMSC) and etc.;
  - Exploring improvement on the multi-omics clustering with the concept of consensus similarity matrix and data-specific similarity matrix 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* 09/2021-08/2022  
**Predicting Patient Adherence with Random Forest**
  - Developing Random Forest base algorithm to explore and predict patient's medication pattern with ECAP medication adherence data.
  - We changed the sampling to Within-cluster sampling and use missing incorporate in attribute in splitting to make random forest compatible with repeated measure incomplete data.
  - We develop Double Robust Random Forest at first as we treated the data as naive data; Followed by Longitudinal

Random Forest as we view the data in longitudinal format; And the latest Alternating Event Random Forest as we see the data in a Markov Process, to find best method for predicting a patient's future medication pattern.

- **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.

## Conference Appearance

- **IHOWD 2024** Portugal  
Oral Presentation 03/2024  
**HIV cohort synthesis data generation**  
– Generate publishable HIV-cohort EHR data with Generative Adversary Network.

## 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