

# Jialiang Hua

New York, NY, 10032 | (646) 346-0432 | jh4323@columbia.edu | github.com/jialianghua

## EDUCATION

**Columbia University, Mailman School of Public Health**  
Master of Science in Biostatistics Theory and Methods Track

**New York, U.S.**  
May 2023

**Soochow University, School of Medicine**  
Bachelor of Medicine (English based, WES certified U.S. MD-equivalent)

**Suzhou, China**  
June 2020

## PROFESSIONAL EXPERIENCE

**Columbia University Department of Psychiatry and New York State Psychiatric Institute**  
*Biostatistician*

**New York, U.S.**  
July 2023 – Present

- Conducted comprehensive data sanitation and statistical analyses for 6 clinical studies (including 3 R01 studies). Communicated findings with doctors and researchers, and co-drafted statistical analysis for various published papers.
- Utilized advanced statistical methods, including generalized nonlinear mixed-effects models and mediation models, to analyze data from longitudinal studies, and contributed to significant findings on the effects of gestational SSRI exposure in the development of functional gastrointestinal disorders.
- Developed complex R and SAS algorithms to sanitize and analyze the Medicaid MAX and TAF data; investigated long-term outcomes of buprenorphine (BUP) treatment in opioid use disorder.
- Led comprehensive statistical analysis for NIH-funded R01 grant; evaluated and validated proton magnetic resonance spectroscopy (<sup>1</sup>H MRS)-based biomarkers for assessing and monitoring schizophrenia; contributed to clinical literature on treatment stratification and early intervention strategies.

**Department of Biostatistics, Columbia University**  
*Research Assistant*

**New York, U.S.**  
April 2022 – April 2023

- Using R, conducted in-depth research of the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) dataset. Ensured data quality through thorough validation and pre-processing, handling missing values, and performed comprehensive statistical analysis.
- Developed predictive models using advanced machine learning techniques. Used cross-validation to train weighted random forest models for identifying key risk factors of mania in people with major depressive disorder. Accelerated the modeling process via parallelization. Evaluated model efficacy and the impact of variable transformation on predictive power.
- Explored various techniques for addressing data imbalance, including balanced random forest, weighted random forest, over-sampling, down-sampling, and Synthetic Minority Over-sampling Technique (SMOTE).
- Produced research abstracts and reports in written and graphical form for seminars and practicum symposium.

**Pfizer Inc.**

*Clinical Data Manager (Wuxi Clinical Outsourced)*

**Shanghai, China**  
October 2020 – April 2021

- Performed data reviews, data reconciliation and query management with SQL and Excel to support 3 ongoing clinical trials.
- Provided over 30 data management metrics reports to study team and stakeholders.
- Conducted user acceptance tests of eCRFs and ensured filing of all related documentation.

## RESEARCH PROJECTS

**Understanding weights options in Estimated Marginal Means.**

*Project Lead, [Github Link](#)*

**New York, U.S.**  
December 2023

- Pioneered research on weights options in Estimated Marginal Means (EMMs), yielding insights through empirical testing and mathematical analysis on the selection of weights for application practice.
- Collaborated with team members to investigate the “type” and “regrid” options in EMMs; derived calculation of EMMs from models with link function for emmeans function in R, Genmod lsmeans and %margins macro in SAS.

**Modeling of Innocent Death Rate Made By Police in the U.S.**

*Project Team Member, [Github Link](#)*

**New York, U.S.**  
September 2021 – December 2021

- Collected, cleaned, and processed raw data of death records on 50 US states from 2010 to 2020.
- Constructed multiple linear regression model to investigate the association between innocent death rate by police and potential factors. Conducted model diagnostics to validate model assumptions.
- Authored the final report and built a website aggregating all the figures, charts and results.

## SKILLS

**Language:** English (professional), Mandarin Chinese (native)

**Computer:** Proficient in R, SAS, SQL, Microsoft Word, Excel, PowerPoint. Experience with Oracle Health Sciences and Python.

**Certifications:** Python for Everybody Specialization by University of Michigan on Coursera.