Zichen Liu

Los Angeles, CA | zichenliu@ucla.edu | [phone # hidden] | zichenliu.me

Education

University of California, Los Angeles, PhD in Biostatistics (in-progress) Sep 2022 – Present

University of Washington, MS in Biostatistics Sep 2020 – Mar 2022

Pomona College, BA in Chemistry Aug 2014 – May 2018

Highlights

Languages: R, Python, SQL, Julia, Java, SAS, Unix/Linux, LaTeX

Skills: Data analysis & visualization, statistical modeling, Bayesian computation, machine learning

Other: Scientific writing (bit.ly/zlpubs), Agile project management (Jira & Asana), code versioning (GitHub)

Experience

Clinical Data Analyst, UCLA Health Jonsson Comprehensive Cancer Center Jul 2024 – Present

- Organized secure clinical data request for 500+ colonoscopy patients enrolled in ongoing NIH/NCI R01 trial
- Extracted and analyzed abnormal screening test data for 2000+ patients from 6 clinics to inform trial design

Intern, Biostatistics Oncology, Gilead Sciences – Foster City, CA Jun 2025 – Aug 2025

- Designed simulation studies in R to evaluate 10+ methods improvements for Bayesian optimal interval phase I/II trial designs; presented results to 50+ biostatisticians in department

Teaching Assistant, UCLA Department of Biostatistics Jan 2024 – Jun 2024

- Held weekly office hours, graded 10+ assignments, and provided project feedback for 30+ students

Graduate Student Researcher, UCLA Department of Biostatistics Jun 2023 – Mar 2024

- Designed 10+ tables/figures for professor's textbook, developed R package to accompany text, edited proofs

Statistical Consultant, Public Health - Seattle & King County – Seattle, WA Oct 2021 – Mar 2022

- Modeled housing instability using survival analysis in R, presented to stakeholders, published in policy journal

Intern, Genentech – South San Francisco, CA Jun 2021 – Sep 2021

- Modeled associations between respiratory disease outcomes and physical activity in R using high-dimensional biobank and wearable technology data

Data Specialist, Institute for Health Metrics and Evaluation – Seattle, WA Aug 2020 – Feb 2021

- Built an efficient Python pipeline to estimate US healthcare cost and utilization at the county/race granularity
- Designed a data warehouse to store incoming clinical data with automatic quality checks
- Standardized and formatted 6 clinical datasets for the Global Burden of Disease Study using Python

Data Analyst, Institute for Health Metrics and Evaluation – Seattle, WA Jun 2018 – July 2020

- Modeled 30+ injuries for the Global Burden of Disease Study using crosswalking (to incorporate a variety of data types) and spatial-temporal smoothing methods in R
- Created 15+ tables/figures, updated methods appendices, and edited proofs for 10+ medical journal publications
- Built a neural network in Python with Keras to detect new data sources from injury-related keywords

Projects

powertools (R package with 4k+ downloads) CRAN.R-project.org/package=powertools

- Developed 50+ functions for power & sample size calculations with robust type-checking and error-catching