Yuanzhi Yu

680 N. Lake Shore · Chicago · IL 60611 · yuanzhi.yu@northwestern.edu · (646)-284-2789

EDUCATION

Northwestern University
Doctor of Philosophy, Biostatistics

Chicago, IL
2022–Present

Columbia University
Master of Science, Biostatistics

New York, NY
2019–2021

Wuhan University
Bachelor of Science, Life Science and Biotechnology

Wuhan, CHN
2015–2019

PROFESSIONAL EXPERIENCE

Linked **in**

Northwestern University, Feinberg School of Medicine
Graduate Researcher

Chicago, IL
Nov 2022 – Present

– Employ deep learning (Transformer & LSTM) and reinforcement learning models (Q-Learning) to address healthcare data challenges, leading to a 45% improvement in model precision and a 30% improvement in recall for gestational diabetes projects.

– Conducted comparative analysis of different imputation methods on healthcare data, enhancing the efficiency by reducing running time by 60% and eliminating 35% of redundant steps in the data processing pipeline for chronic kidney disease research projects.

– Utilized R for complex data management and PyTorch for implementing machine learning models, and leveraged extensive EMR data for optimizing care interventions, resulting in a 15% increase in patient care efficiency.

– Lead bi-weekly interdisciplinary knowledge-sharing sessions for physicians and medical students with emphasis on integration of statistical and machine learning applications in healthcare, fostering collaborative learning and team synergy.

Columbia University, Mailman School of Public Health
Data Analyst & Research Assistant

New York, NY
Apr 2020 – Jul 2022

– Innovated imputation methods to mitigate measurement errors for multivariate environmental health data, resulting in a 60% reduction in data inconsistencies. The method was subsequently adopted as the standard by the research team.

– Utilized Random Forest, XGboost and Mixed Effects Logistic Regression models for targeted metabolomics analysis in a case-control study, successfully identifying 5 promising biomarkers from a pool of 17680 candidates, marking a significant progress for subsequent investigations.

– Employed regression, Bayesian, and enrichment analyses to investigate associations between plasma metabolomics assays and Chronic Fatigue Syndrome, contributing to a published paper in a peer-reviewed journal.

– Served as the Lead Teaching Assistant for statistical courses, where I spearheaded a team of teaching assistants, orchestrated lecture sessions in collaboration with professors, and conducted tutorial sessions. My efforts significantly enhanced the educational experience for students.

Pfizer
Quality Control, Intern

Hangzhou, CHN
Jan 2019 – Apr 2019

– Established a one-click functionality for data management, storage, and visualization, significantly enhancing processing speed and efficiency.

– Coordinated with cross-functional teams to implement data-driven models for quality control, leading to a 15% improvement in defect detection.

Technical Skills

- **Programming Languages:** R, Python, R Shiny, HTML, C++, JAVA, SQL, SAS.
- **Technologies and Tools:** Git, LaTeX, Docker, Linux (Usage and Shell Scripting).
- **Machine Learning and Statistical Analysis:** Missing Data Imputation, Bayesian Analysis, Deep Learning, Reinforcement Learning, Time Series Analysis, Causal Inference, A/B testing, Experimental Design, Simulation.

Publications

Google Scholar

Peer-Reviewed Publications

- **Yuanzhi Yu**, Roderick J. Little and Qixuan Chen. Multiple Imputation of More than One Environmental Exposure with Non-differential Measurement Error. *Biostatistics*, kxad011, 2023.
- Xiaoyu Che, Christopher R. Brydges, **Yuanzhi Yu**, et al. Metabolomic Evidence for Peroxisomal Dysfunction in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *International journal of molecular sciences*, 23(14), 7906, 2022.
- Filippo Ravalli, **Yuanzhi Yu**, Ana Navas-Acien, et al. Uranium and metal concentrations in community water systems across the United States, 2006-2011. *The Lancet. Planetary health*, 6(4), e320–e330, 2022.
- Zhanning Yang, **Yuanzhi Yu**, and Fuxu Cui. Toxic effects of single-armed carbon nanotubes on Pacific Crassostrea gigas. *Asian Journal of Ecotoxicology*, (01), 90-98, 2019.
- **Yuanzhi Yu**, Wei Zou. Progress in Molecular Mechanism of Metabolic Related Signal Transduction in Type II Diabetes. *Biotechnology Letters*, (04),564-570, 2018.

Submitted & Under Review Publications

- **Yuanzhi Yu**, Yiming Li, Yuan Luo. Evaluation of Imputation Strategies in Medical Time-Series Data: Insights from the MIMIC-III Data. Submitted to *Statistics in Medicine.*
- Seonghun Lee, Guohua Li, Stanford T. Chihuri, **Yuanzhi Yu**, Qixuan Chen. Using Data Fusion and Multiple Imputation to Correct for Misclassification in Self-reported Exposure: A Case-Control Study of Cannabis Use and Homicide Victimization. Under revision in *Epidemiology.*
- Yiming Li, **Yuanzhi Yu**, Sanaz Ghotbaldini, Yuan Luo. Metabolomic Insights into Aging and Alzheimer's Disease: Biomarkers, Therapeutic Targets, and Multi-Omics Advances. Submitted to *Briefings in Bioinformatics.*
- Domonique M. Reed, **Yuanzhi Yu**, Jessica E. Justman, Qixuan Chen. Integrative Data Analysis for Harmonizing Measures from Multiple Data Sources: Assessing Population-level Viral. Submitted to *Epidemiology.*

Projects

GitHub

- Development and Maintenance of `MI.Lasso` R Package: An implementation for multiple imputation in datasets with missing values, available at `https://github.com/yy3019/MI.lasso`. Over 100 downloads on GitHub.
- Public Drinking Water Contaminant Exposure Estimate Dashboard: An interactive dashboard for public awareness on drinking water quality, available at `https://msph.shinyapps.io/drinking-water-dashboard`. Garnered over 100,000 views and featured in a NYC health newsletter.
- NYC Neighborhood COVID-19 Dashboard: Collaborated on a project to track COVID-19 statistics across NYC neighborhoods using R Shiny, available at `https://msph.shinyapps.io/nyc-neighborhoods-covid`. Over 25,000 views and shared by different health departments.
- Exploring the Fragility Index and Clinical Trials: Web-based project to visualize and analyze the fragility index in clinical trials, available at `https://yy3019.github.io/P8105_final_project.github.io`.

Miscellaneous

- Student Paper Award Winner, *Association of Clinical and Translational Statisticians*, 2023.
- Biostatistics In Action: Best Dashboard, *Columbia University*, 2021.
- National Scholarship & Merit Student & Excellent Student Leader , *Wuhan University*, 2019.
- Bronze Medal, *National Youth Science and Technology Innovation Competition*, 2015.