

Ruishan Lin

📍 Fairfax, VA 📩 rlin8@gmu.edu ☎ +1 646-578-9222 💬 rs-lin 💬 ruishanlin

Education

George Mason University <i>PhD Candidate in Statistical Science (Passed Qual. Exams)</i>	<i>Aug 2021 – May 2026</i>
New York University <i>BA in Mathematics, Minor in German</i>	<i>GPA: 3.77/4.0</i> <i>Sept 2016 – May 2020</i> <i>GPA: 3.74/4.0 (cum laude)</i>

Work Experience

Biomarker Statistician Intern <i>Sanofi</i>	<i>Cambridge, MA</i> <i>May 2025 – Aug 2025</i>
Project Name: Identifying Influential Substructures in Atopic Dermatitis (AD) Disease Network for Understanding Drug Response Mechanisms.	
Supervisors: <i>Drs. Siying Huang and Wenting Wang</i>	

- Applied graph-based deep learning techniques, including Graph Convolutional Networks (GCN), Graph Attention Networks (GAT), and the Graph Information Bottleneck framework (GIB), using Python and PyTorch.
- Constructed patient-specific disease networks from transcriptomic data (responders vs. non-responders) using the LIONESS framework to enable individualized network analysis.
- Developed and validated classification models using GCN, GAT, and GIB to predict drug response in Atopic Dermatitis (AD) patients, achieving 99% accuracy on simulated datasets.
- Identified key substructures within AD gene networks that are responsible for the drug's mechanism of action, revealing biologically meaningful patterns and informing potential drug target discovery.

Clinical Data Science Intern <i>Boehringer Ingelheim</i>	<i>Ridgefield, CT</i> <i>May 2024 – Aug 2024</i>
Project Name: PRO-SAFE: A Bayesian Model for Predicting Safety Signals in Clinical Trials.	
Supervisors: <i>Drs. Dooti Roy, Lisa Neums, and Arnab Maity</i>	

- Leveraged historical clinical trial data to inform and improve the efficiency of safety signal detection using Bayesian Hierarchical Models with nested indications and trials.
- Designed various scenarios to assess model prediction accuracy and performance under varying heterogeneity levels across arms and covariates.
- Greatly reduced the simulation time by implementing parallel computing via the *targets* framework in *R*.
- Presented the project at the **2024 ASA Biopharmaceutical Section Regulatory- Industry Statistics Workshop**, demonstrating the project's potential to inform future clinical trial safety predictions.

Graduate Research and Teaching Assistant <i>George Mason University</i>	<i>Fairfax, VA</i> <i>Aug 2021 – Dec 2025</i>

- Spearheaded three interdisciplinary research projects with a focus on Phenology and Biology, collaborating closely with advisors and field experts.
- Executed large-scale simulations in R on high-performance computing servers, supporting complex data analysis for research initiatives.
- Assisted in teaching four undergraduate and graduate-level statistics courses, handling grading, hosting weekly office hours, and responding to course-related inquiries.
- Courses Assisted: STAT 250 - Introductory Statistics I, STAT 344 - Probability and Statistics for Engineers and Scientists I, STAT 354: Probability and Statistics for Engineers and Scientists II, STAT 515 - Applied Statistics and Visualization for Analytics.

Research Projects

Detection of Multiple Change Points in Non-Stationary Network Autoregression Models

2024 - present

Supervisor: Dr. Abolfazl Safikhani

- Advanced the [Network Autoregression \(NAR\) Model](#) by exploring new techniques for detecting change points within network-dependent time series data, focusing on changes in model parameters to enhance interpretability.
- Proposed a novel rolling-window approach to detect change points efficiently and easily adapt to evolving network structures and parameter variations.
- Illustrated the models effectiveness by recovering seizure times on an electroencephalogram (EEG) brain scan dataset.
- (*Under Review*) **Lin, R. and Safikhani, A.** *Multiple Change Point Detection for Non-Stationary Network Auto-Regressive Models with Node-Level Dynamics.*
- Tools Used: *R, ITSM*

A Nonparametric Bayesian Model to Adjust for Monitoring Bias

2023 - present

Supervisors: Drs. Jonathan Auerbach and David Kepplinger

- Developed a Bayesian model incorporating Penalized B-splines to correct for monitoring bias in citizen science data and copulas for leveraging information across spatial-domain, enabling more accurate estimation of event timing in environmental studies.
- Applied this model to detect environmental stress indicators linked to climate change by analyzing lilac blooming patterns.
- Presented the project on **Joint Statistical Meetings 2024**, demonstrating the model's potential for ecological monitoring applications.
- (*Under Review*) **Auerbach, J., Crimmins, T. M., Kepplinger, D., Lin, R., Wolkovich, E. M.** *A Nonparametric Bayesian Model to Adjust for Monitoring Bias with an Application to Identifying Environments Stressed by Climate Change.* [\[arXiv Link\]](#)
- Tools Used: *R, Stan*

Course Projects

Interactive Dashboard of US Population Census Data (2011–2023)

[R Shiny Dashboard](#)

- Designed an interactive dashboard for visualizing U.S. population trends, migration patterns, and racial demographics over time using dynamic maps, stacked bar charts, and time series plots.
- Tools Used: *R Shiny, R, Plotly, Leaflet*

Handling Missing Data for Phase III Clinical Trials

[R Shiny Dashboard](#)

- Conducted extensive simulations to evaluate the effectiveness of various imputation methods under different missing data patterns, assessing their impact on clinical trial outcomes.
- Presented findings at the **Statistics in Pharmaceuticals (SIP 2023) Conference**, highlighting practical recommendations for handling missing data.
- Tools Used: *R Shiny, R, SAS*

Time Series Forecasting with Deep Learning Methods

[View Source on GitHub](#)

- Conducted a comparative analysis of deep learning models (LSTM and Transformer) versus traditional statistical models (Seasonal ARIMA) for forecasting temperature data, using RMSE to measure accuracy.
- Tools Used: *Python, PyTorch, R, ITSM*

Technologies

Programming Skills: **R (Shiny)**, MATLAB, SAS, **Python/PyTorch**, HTML, Tableau, Parallel Computing

Quantitative Skills: **Bayesian Methods**, Categorical Data Analysis, **Deep Learning**, **Graph / Network Analysis**, Group-Sequential Design, Longitudinal Data Analysis, **Machine Learning**, Nonparametric Statistics, Optimization, Regression Methods, **Time Series**, **Visualization**.

Leadership and Teamwork Experience

Vice President, Statistics Graduate Student Association
Site Ambassador, New York University Berlin Campus

2021 - 2025
2019

Awards

First Place: Statistical Significance Poster Competition ↗

Joint Statistical Meetings (JSM) 2025

PhD Student Poster Competition Award

StatConnect 2025

Distinguished Service and Leadership Award

May 2025

Department of Statistics, George Mason University