

Ruishan Lin

📍 Fairfax, VA ✉ rlin8@gmu.edu ☎ +1 646-578-9222 🌐 rs-lin in ruishanlin

Education

George Mason University

PhD Candidate in Statistical Science (Passed Qual. Exams)

Aug 2021 – May 2026

GPA: 3.77/4.0

New York University

BA in Mathematics, Minor in German

Sept 2016 – May 2020

GPA: 3.74/4.0 (cum laude)

Work Experience

Biomarker Statistician Intern

Sanofi

Cambridge, MA

May 2025 – Aug 2025

Project Name: **Identifying Influential Substructures in Atopic Dermatitis (AD) Disease Network for Understanding Drug Response Mechanisms.**

Supervisors: Drs. Siying Huang and Wenting Wang

- Applied graph-based deep learning techniques, including Graph Convolutional Networks (GCNs), Graph Attention Networks (GATs), and the Graph Information Bottleneck (GIB) framework, 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 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

Boehringer Ingelheim

Ridgefield, CT

May 2024 – Aug 2024

Project Name: **PRO-SAFE: A Bayesian Model for Predicting Safety Signals in Clinical Trials.**

Supervisors: Drs. Dooti Roy, Lisa Neums, and Arnab Maity

- 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.
- Reduced the time to run simulations significantly by implementing parallel computing using the *targets* framework in R.
- Presented the methodology and the results 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

George Mason University

Fairfax, VA

Aug 2021 – Present

- 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.

Machine Learning Engineer Intern

HiThink Financial Services Inc.

New York, NY

Feb 2020 – Apr 2021

- Engineered chatbots that can answer users' questions about real-time financial market information and provide interactive visualizations using *PyTorch* and *TensorFlow*.
- Built and maintained data pipelines for financial data through continuous web scraping, transformation, and cleaning processes using *Python*, streamlining data preparation.

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 rolling-window approach to detect change points efficiently and easily adapt to evolving network structures and parameter variations.
- Conducted inferences on the model's performance, and illustrated the models effectiveness by recovering seizures times on an electroencephalogram brain scan dataset.
- 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.
- (Manuscript) 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*

Reassessing the Growing Degree Day Model

2023 - present

- Enhanced the traditional Growing Degree Day Model by integrating it with statistically robust activation functions, aiming to improve predictive accuracy for phenological events.
- Conducted comprehensive simulations to test the model's performance, and validated results through cross-validation and model stacking with real-world datasets.
- Presented the project on the **ICORS meets DSSV 2024** conference.
- (Manuscript) **Lin, R.**, Kepplinger, D., Auerbach, J. A Statistical Reassessment of the Growing Degree Day Model for Predicting Phenological Events
- Tools Used: *R*, *Parallel Computing*

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 - present
2019

Awards

PhD Student Poster Competition Award

March 2025

StatConnect 2025

Distinguished Service and Leadership Award

May 2025

Department of Statistics, George Mason University