

Runqiu(Rachel) Wang

Cell: +1(202)7107922 | Email: runqiurachelwang@gmail.com | [LinkedIn](#) | [Google Scholar](#) | [GitHub](#)

SUMMARY

Ph.D. candidate in Biostatistics with 2+ years of industry experience in clinical trials and drug development. Strong background in clinical trials, SAP Development, and FDA submissions. Research interests in Bayesian methods, multiple testing, high-dimensional inference, and machine learning, with a strong track record in developing statistical methods for complex healthcare data. Successfully applied statistical innovations to multi-center EHR data, brain image data and GWAS studies. Proficient in both theoretical research and practical implementation in pharmaceutical development. Strong communicator with 5+ years of biostatistical consulting experience across diverse teams.

EDUCATION

University of Nebraska Medical Center <i>Ph.D. in Biostatistics, GPA: 3.92/4.00</i> Thesis: <i>Novel methods for False Discovery Rate control in high-dimensional variable selection.</i>	Omaha, NE <i>August 2021- Anticipated Summer 2025</i>
Georgetown University <i>M.S. in Biostatistics, GPA: 3.88/4.00</i> Thesis: <i>Predictors of time-to-event outcomes in the ALLHAT trial identified with machine learning.</i>	Washington, DC <i>August 2017 - December 2018</i>
Hong Kong Baptist University <i>B.S. in Statistics, GPA: 3.30/4.00</i>	Zhuhai, China <i>September 2013 - June 2017</i>

TECHNICAL SKILLS

Certifications:	SAS Certified Advanced Programmer; SAS Certified Based Programmer
Programming Languages:	R, SAS, Python, MATLAB, C++
Tools:	R shiny, MySQL, UNIX commands for HPC, Parallel computing

WORKING EXPERIENCE

U.S. Food and Drug Administration (CDER) ORISE Fellow (Intern)	Silver Spring, Maryland <i>May 2024- August 2024</i>
---	--

*Working on a research project to investigate the **missing data** in Ambulatory Blood Pressure Monitoring (ABPM) study.*

- Investigated patterns of missingness in Ambulatory Blood Pressure Monitoring (ABPM) studies and evaluated statistical methods including Multiple Imputation (MI), Inverse Probability Weighting (IPW), and Augmented Inverse Probability Weighting (AIPW).
- Conducted a simulation study to evaluate and compare the performance of MI, IPW and AIPW under two missing data scenarios: Missing Completely at Random (**MCAR**) and Missing at Random (**MAR**).
- Developed evidence-based recommendations for ABPM study designs by identifying optimal statistical methods based on specific missing data patterns and comprehensive simulation results.

Firma Clinical Research, LLC. Statistician	Hunt Valley, Maryland <i>January 2021–July 2021</i>
---	---

- Executed statistical activities under supervision for clinical trials, including **CRF** review, **SAP** development, statistical analysis accuracy validation, and results interpretation, particularly in Phase **II/III** studies with Integrated Summary of Safety (**ISS**) and Integrated Summary of Efficacy (**ISE**) submissions.
- Acted as a primary statistical contact for **renal** diseases and **eye** diseases studies, assisting in the design of clinical trials, sample size justification, randomization scheme, and interim analysis for **Phase II/III** Clinical Trials.

Associated Statistical Programmer	<i>June 2019– January 2021</i>
--	--------------------------------

- Specialized in creating **CDISC**-compliant SDTM and ADaM datasets, along with standard tables, figures, and listings (**TFLs**) for Phase II/III clinical trials.
- Supported FDA submission processes by preparing **Define XML** files, as well as Study Data Reviewer's Guide (**SDRG**) and Analysis Data Reviewer's Guide (**ADRG**).

Johnson & Johnson Statistician (Intern)	Beijing, China <i>May 2018-August 2018</i>
--	--

Working on a research project to predict the probability of drug rivaroxaban's adverse events in cardiovascular diseases.

- Processed and integrated cardiovascular study data using SQL, developing SAS macros for data cleaning, labeling, and descriptive analysis.
- Used machine learning algorithms such as lasso, naive Bayes classifier, classification tree, random forests, and gradient boosting machine using **Python** to predict the probability of rivaroxaban's adverse events (mainly bleeding events) and their associated risk factors.
- Achieved prediction rates at least 85%, and effectively visualized, and reported the results using **R shiny**.

RESEARCH EXPERIENCE

University of Nebraska Medical Center

Omaha, Nebraska

Graduate Research Assistant

August 2021–Present

Project 1: Bayesian Logistic Regression for Heart Failure Survival Prediction

- Developed and implemented a Bayesian logistic regression model with normal priors to predict heart failure survival using a latent variable approach.
- Identified key clinical predictors (age, ejection fraction, serum creatinine) and validated model convergence through trace plots and auto-correlation analysis.
- Compared results with frequentist methods, confirming consistency while enhancing interpretability through Bayesian inference.

Project 2: Machine Learning Models Decoding the Association Between Urinary Stone Diseases and Metabolic Urinary Profiles

- Designed and implemented **feature engineering** techniques to optimize biomarker selection for disease classification.
- Applied **supervised** and **unsupervised learning algorithms** to detect complex metabolic patterns predictive of disease risk with more than **90%** accuracy.
- Used **deep learning algorithms** such as Convolutional Neural Networks (**CNNs**) and Graph Neural Networks (**GNNs**) to analyze biomedical imaging data related to metabolic pathways.

Project 3: Controlling FDR in Selecting Group-level Simultaneous Signals

- Created a knockoff-based algorithm for identifying mutual signals from group-level conditional independence tests, ensuring **exact false discovery rate (FDR) control** under finite sample settings while achieving **over 90% statistical power**.
- Contributed to the National COVID Collaborative Cohort (N3C) project by addressing heterogeneity and missingness in large-scale multi-site **EHR** datasets.

Project 4: Mediation Selection with FDR Control for Nonlinear Models

- Developed a novel algorithm for causal mediation selection, ensuring FDR control in high-dimensional **nonlinear** mediation models where traditional methods fail.
- Implemented the method to identify structural MRI mediators between marriage status and dementia progression in the ADNI study.

Georgetown University

Washington, DC

Student Research Assistant

January 2018-May 2018

- Used **Plink** to grasp raw data in Framingham Heart Study of Genetic Analysis Workshop 16 to preprocess the data.
- Contributed to developing stochastic functional linear models to analyze the cleaned dataset at gene levels and to measure quantitative traits **longitudinally**.

Hong Kong Baptist University

Zhuhai, China

Student Research Assistant

August 2016-July 2017

- Implemented Normalized Iterative Hard Thresholding (NIHT) Algorithm in **R** to compare with other variables selection methods, including Lasso, Smoothly Clipped Absolute Deviation (SCAD), and Minmax Concave Penalty (MCP).
- Used simulation studies to show the interpretability and prediction accuracy of NIHT Algorithm in logistic model.
- Applied data in genome-wide association study (**GWAS**) to identify associated taxa of bacteria for obesity via NIHT Algorithm.

PUBLICATIONS (SELECTED)

Statistical Methodology

- Wang R., Dai, R., Dai, H., & Zheng, C. (2023). Controlling FDR in selecting group-level simultaneous signals from multiple data sources with application to the National Covid Collaborative Cohort data. *arXiv preprint arXiv: 2303.01599*.
- Wang R., Dai, R., Huang, Y., Neuhaus, M., Lampe, J., Raftery, D., Tabung, F., & Zheng, C. (2023). Variable selection with FDR control for noisy data – an application to screening metabolites that are associated with breast and colorectal cancer. *arXiv preprint arXiv: 2310.06696*.
- Wang, R., & Peng, X. (2022). Normalized Iterative Hard Thresholding in High Dimensional Logistic Regression With Metagenomics Data. *International Journal of Intelligent Technologies & Applied Statistics*, 15(1).
- Zhang, B., Wang, S., Mei, X., Han, Y., Wang, R., Fang, H., Chiu, C., Ding, J., Wang, Z., Wilson, A., Bailey-Wilson, J., Xiong, M., and Fan, R. (2022). Stochastic function linear models for gene-based association analysis of quantitative traits in longitudinal studies. *Statistics and its interface*. 15.

Statistical Application

- Ma, L., Qiao, Y., Wang, R., Chen, H., Liu, G., Xiao, H., & Dai, R. (2024). Machine Learning Models Decoding the Association Between Urinary Stone Diseases and Metabolic Urinary Profiles. *Metabolites*, 14(12), 674.
- Vieira, H. M., Kasper, D. P., Wang, R., Smith, L. M., Enke, C. A., Bergan, R. C., ... & Baine, M. J. (2023). Comparison of sequential versus concurrent chemoradiation regimens in non-metastatic muscle-invasive bladder cancer. *Radiation Oncology Journal*, 41(3), 154-162.
- Lu, Y., Wang R., Norman, J., & Yu P. (2023). Loneliness trajectory and risk of cardiovascular disease among middle-aged and older adults. *Nutrition, Metabolism and Cardiovascular*.
<https://doi.org/10.1016/j.numecd.2023.10.024>.

PRESENTATIONS

- Wang R., Huang D., Weng, Y. & Zhao, X. (2024). Missing Data in Ambulatory Blood Pressure Monitoring (ABPM) Studies. OB Science Day, FDA, September 12, 2024 (Poster).
- Wang R., Dai, R., Huang, Y., Neuhaus, M., Lampe, J., Raftery, D., Tabung, F., & Zheng, C. (2023). Variable selection with FDR control for noisy data – an application to screening metabolites that are associated with breast and colorectal cancer. 2024 JSM, Portland, OR, August 8, 2024 (Top-Contributed Paper Presentation).
- Wang R., Dai, R., Dai, H., & Zheng, C. (2024). Controlling FDR in selecting group-level simultaneous signals from multiple data sources with application to the National Covid Collaborative Cohort data. 2024 ENAR, Baltimore, MD, March 12, 2024 (Oral-Contributed Talk).
- Wang R., Colin O. Wu (2018). Identifying Risk Factors Associated with Myocardial Infarction, Stroke and Cardiovascular Death Using Machine Learning Techniques based on BARI 2D Data. 2018 Student Poster Presentation, Georgetown University, December 12, 2018 (Poster).

HONORS and AWARDS

- | | |
|--|-----------|
| • Dean's List, Bachelor of Science, Statistics, Hong Kong Baptist University, China | 2013-2017 |
| • Third Prize in 2016 Contemporary Undergraduate Mathematical Contest in Modeling, China | 2016-2017 |

PROFESSIONAL MEMBERSHIPS

American Statistical Association (ASA) Member	2022-Present
Eastern North American Region (ENAR) Member	2022-Present
International Biometric Society (IBS) Member	2022-Present