

Miaolei Bao

Email: baomiaolei@gmail.com | Phone: 206 710 8016 | Website | LinkedIn | Github

Summary of Qualifications

Master-level Biostatistician with experience in clinical and healthcare data and statistical modeling, offering analytical skills in programming (R, Python, SAS, SQL). Strong problem-solving and collaboration abilities; excellent written, verbal, and visual communication skills, detail-oriented, with demonstrated success in working on complex datasets.

Skills

Technologies: R (ggplot2, dplyr, tidyverse), Python (pandas), SAS, Github, Latex, R Markdown, SQL, Microsoft Office

Statistical: Linear Regression, Logistic Regression, Nonparametric Analysis, Time Series Analysis, Generalized Linear Model, Machine Learning, Sample Size Calculation, Hypothesis Testing, Categorical Data, Bayesian Inference, Simulation, Survey Data

Work Experience

Research Assistant, Fred Hutchinson Cancer Center | Seattle, WA 02/2024 – Present

- Prepared **statistical analysis plans** and presentations under the supervision of senior biostatisticians for **manuscripts**.
- Cleaned data from databases and **case report forms** using **R**, and generated datasets, tables, and figures to communicate findings and support **longitudinal data** analysis using **Generalized Linear Models** and **Generalized Estimating Equations**.

Research Experience

Valley View Health Center, Pharmacist-led Diabetes Intervention Evaluation | Seattle, WA 09/2024 – 03/2025

- Cleaned and standardized **38,000+ Electronic Health Records (EHRs)** collected across **11** health center sites, imputed **10% missing data** and performed quality checks, ensuring high data quality and readiness for statistical analysis.
- Designed and implemented **statistical analysis plans** in R, utilizing **linear mixed effect models**, **Kaplan-Meier curves**, and **Cox proportional hazards models** to assess the effectiveness of a pharmacist-led intervention for diabetes management.
- Conducted meetings with medical professionals, translating statistical findings into clinical insights in non-technical language.

Health Metrics Prediction Using Demographic and Health Surveys | Seattle, WA  05/2024 – 03/2025

- Built pipelines to automate the extraction and modification of **4,000+** survey metric preprocessing functions using **Python**.
- Utilized **Natural Language Processing (NLP)** to match function labels with corresponding formal variable descriptions.
- Wrote documentation and applied Bayesian hierarchical models in **R** to generate national and regional estimates.
- Contributed to an **R package**, assisted in resolving identified technical issues, prepared **reports** and created various **visualizations** for group meetings with academic collaborators, government representatives, and United Nations experts.

Thorax Disease Prediction via Machine Learning-based X-ray Classification | Seattle, WA  09/2024 – 12/2024

- Developed **prediction models** using **Convolutional Neural Networks** to classify thorax disease images using chest X-ray data with over 5,600 images, employed **F1 score** to ensure robustness, and achieved an accuracy of **93.3%**.
- Utilized **TensorFlow** and **Keras** for model architecture development, applied dropout techniques to prevent overfitting and improve model generalization, and **optimized** hyperparameters such as batch sizes for high-dimensional image data.

Statistician, Randomized Prospective Clinical Trial | Seattle, WA 04/2024 – 06/2024

- Collaborated with a multi-disciplinary team to formulate the experiment design and led the drafting of the **clinical trial protocol**, employing a factorial design to evaluate the effects of two behavioral interventions on targeted outcomes.
- Specified expected treatment effects, calculated the required trial **sample size**, defined the **randomization** and blinding strategies, outlined methods for handling missing data, and detailed the approaches for primary and secondary analyses.

Genomic Analysis of Tobacco for Genes Related to Chemical Constituents | Hangzhou, China 10/2022 – 06/2023

- Calculated **descriptive statistics**, performed correlation analysis and conducted **linear regression** and **analysis of variance** on the concentrations of commercially important tobacco chemicals using **SAS**.
- Identified significant gene-gene and gene-environment interaction effects using the gene linkage map and predicted 21 candidate genes useful for tobacco breeding using protein databases.

Education

University of Washington, MS in Biostatistics (Data Science), GPA: 3.95/4.00 | Seattle, WA 09/2023 – 03/2025

Zhejiang University, BS in Ecology, GPA: 3.95/4.00 | Hangzhou, China 09/2019 – 06/2023