Ibrahim Hossain Sajal

Biostatistician

🏠 Rockville, MD
☎ (682)-232-8791
@ sajali2@nih.gov
🌐 ihsajal.github.io

About Me

I am a biostatistician with a PhD in Data Science & Statistics, specializing in scalable Bayesian and genetic modeling to drive biomarker discovery and risk prediction in biomedical research.

Education

Ph.D. in Data Science & Statistics
University of Texas at Dallas
Richardson, TX, USA (2019 - 2024)

Master's. in Applied Statistics
University of Dhaka
Dhaka, Bangladesh (2017 - 2018)

Bachelor's in Applied Statistics
University of Dhaka
Dhaka, Bangladesh (2013 - 2016)

Skills

Programming & Computing:
R, Python, SAS, AWS, HPC, Shell scripting

Statistical Modeling:
Survival Analysis, Conditional & Penalized Regression (LASSO), Variable Selection, Cross-Validation, Multiple Imputation, Bayesian Inference, MCMC, Latent Variable Models

Machine Learning:
Supervised & Unsupervised Learning, Simulation Studies

Genomics & Causal Analysis:
GWAS, Mendelian Randomization, Instrumental Variable, Mediation Analysis, Differential Gene Expression

Awards

- Fellows Award for Research Excellence
- Mei Lein Fellowship
- Boyd Harshburger Travel Award
- PhD Research Small Grant Program
- SISG Scholarship
- Academic Excellence Award

Experience

Postdoctoral Fellow
National Cancer Institute, NIH (2025 - Present)

- Identified proteomic biomarkers linking risk factors to cancer using integrative omics and gene expression analyses.
- Developed a scalable, summary-statistics–based multivariate testing framework to improve detection of genetic associations.
- Analyzed pan-cancer phenotypes and protein networks to uncover pathway-level signals relevant to target discovery (KEGG, Hallmark).
- Conducted survival analyses in lung cancer to evaluate prognostic impact of clinical risk factors and polygenic risk scores.

Graduate Research Assistant
The University of Texas at Dallas (2020 - 2024)

- Developed **CBCRisk-Black** and **CBCRisk-Mastectomy**, two personalized absolute risk prediction tools to support clinical decision-making in contralateral breast cancer.
- Analyzed large-scale population and registry datasets (BCSC, SEER) to identify risk factors and estimate cancer incidence and hazard rates.
- Designed haplotype-based genetic association test, **bivariate Quantitative Bayesian LASSO**, for correlated disease phenotypes.
- Extending Bayesian LASSO models to high-dimensional, multivariate settings using mean-field variational inference to enable scalable genetic discovery.

Publications

- **Sajal, I. H.**, & Biswas, S. (2023). Bivariate quantitative Bayesian LASSO for detecting association of rare haplotypes with two correlated continuous phenotypes. Frontiers in Genetics, 14, 1104727.
- **Sajal, I. H.**, Chowdhury, M., Wang, T., Euhus, D., Choudhary, P. K., & Biswas, S. (2022). CBCRisk-Black: a personalized contralateral breast cancer risk prediction model for black women. Breast cancer research and treatment, 194(1), 179–186.

Presentations

- Joint Statistical Meetings 2025 — Invited Talk
- NIH Research Festival — Poster Presentation
- Advances in Statistical and Computational Methods for Analysis of Biomedical, Genetic, and Omics Data — Poster Presentation
- Southern Regional Council on Statistics — Poster Presentation