Ibrahim Hossain Sajal

Email : ihossain2@isrt.ac.bd
Mobile : +1-682-232-8791
LinkedIn : IHSajal
Website : ihsajal.github.io

ABOUT ME

I am a statistician with a PhD in Data Science & Statistics, with 5+ years of experience in statistical genetics and scalable Bayesian modeling to drive biomarker discovery and cancer risk prediction in biomedical research.

EDUCATION

PhD in Data Science & Statistics
School of Natural Science and Mathematics, University of Texas at Dallas   *Aug. 2019 – Dec. 2024*

Masters in Applied Statistics
Institute of Statistical Research and Training, University of Dhaka   *Jan. 2017 – Dec. 2017*

Bachelors in Applied Statistics
Institute of Statistical Research and Training, University of Dhaka   *Jan. 2013 – Dec. 2016*

PROFESSIONAL EXPERIENCE

Postdoctoral Fellow   Jan. 2025 – Present
Division of Cancer Epidemiology & Genetics, National Cancer Institute, National Institutes of Health

- Investigate proteomic intermediaries linking risk factors to renal cancer using two-stage Mendelian Randomization (MR), complemented by mediation analysis, cis-MR, colocalization, and differential gene expression.
- Developed and implemented pipelines for a summary-statistics based multivariate test by aggregating multiple tests with complementary strength in capturing genetic association.
- Conduct multivariate analyses to identify associations with pan-cancer phenotypes, protein modules (WGCNA), and predefined pathways from KEGG and Hallmark.
- Identify pan-caner associations with protein modules using Bayesian Non-Negative Matrix Factorization (BNMF) and ClustImpute.
- Perform survival analyses in RCC cases and examine relationships with risk factors and polygenic risk scores.

Graduate Researcher   Aug. 2019 – Dec. 2024
School of Natural Science and Mathematics, University of Texas at Dallas

- **Genetic association test - bivariate QBL**: Detecting common and rare genetic association with two correlated phenotypes.
  - ∗ Applied Bayesian LASSO to regularize regression coefficients, enhancing the detection of associated haplotypes by leveraging a latent variable to model the correlation between two phenotypes.
  - ∗ Estimated the posterior distribution using MCMC and calculated Bayes factors to identify haplotype effects, achieving significant performance improvements in bivariate QBL over the existing association test, Haplo.score.
  - ∗ Conducted extensive simulations and analyzed GAW 19 exome sequencing data to uncover rare haplotypes associated with systolic and diastolic blood pressures.
- **Multivariate genetic association test**:
  - ∗ Developing a multivariate version of the bivariate QBL test using an efficient machine learning tool, Variational Inference, for approximation of the parameter distribution.
  - ∗ Investigating the statistical properties of the test using simulations and comparing its power to that of bivariate QBL.
- **CBCRisk-Black - Contralateral breast cancer (CBC) risk prediction model for black women.**
  - ∗ Developed a relative risk model for CBC using matched LASSO logistic regression on BCSC data, and performed survival analysis on SEER data to estimate absolute risks.
  - ∗ Validated the CBCRisk-Black model through six-fold cross-validation, achieving a 10% increase in the area under the ROC curve compared to the existing CBCRisk model for 3- and 5-year predictions.
  - ∗ Implemented multivariate imputation by chained equations (MICE) to ensure robustness of CBCRisk-Black and created an R package for accessibility by patients, clinicians, and researchers.
- **CBCRisk-Mastectomy - CBC risk prediction tool for contralateral prophylactic mastectomy (CPM) decision making.**
  - ∗ Developed CBC risk prediction model using only the pre-surgical (mastectomy) risk factors. With risk estimates from this model, a BC patient who is about to undergo mastectomy can decide whether to undergo CPM at the same time.

Technical Skills

Data Analysis: R, Python, SAS, STATA, MATLAB   **Programming Language**: C++

Data management: SQL, Linux commands, HPC environment, Bash scripts, AWS

Documentation: Latex, Microsoft Word, Excel, PowerPoint   **Reference management**: Endnote, Mendeley

Scientific Presentations

- **Multivariate QBL for Detecting Rare Haplotype Association with Correlated Phenotypes using Variational Bayes approach**.
  Joint Statistical Meetings (JSM), Nashville, TN   Aug. 2025.

Poster Presentations

- **Integrative analysis identifies potential proteomic intermediates associated with renal cell carcinoma and its risk factors**.
  National Institutes of Health Research Fair, Bethesda, MD   Sep. 2025.

- **Bivariate QBL for Detecting Rare Haplotype Association with Two Correlated Phenotypes**.
  Advances in Statistical and Computational Methods for
  Analysis of Biomedical, Genetic, and Omics Data (ABGOD), Dallas, TX   Mar. 2023.
  Southern Regional Council on Statistics (SRCOS), Jekyll Island, GA   Sep. 2022.

Publications

1. **IH. Sajal**, A. Song, K. Brown, M. Machiella, P. Kraft, S. Chanock, M. Purdue, and D. Dutta (2026) *Integrative analysis prioritizes proteins associated to renal cell carcinoma and its risk factors*, **Currently under reivew**.

2. **IH. Sajal**, R. Pfeiffer, I. Jatoi, M. Gail, R. Cecchini, P. Choudhary, and S. Biswas, (2026) *CBCRisk-Mastectomy: A Risk Prediction Tool for Contralateral Prophylactic Mastectomy Decision Making*, **Currently under reivew**.

3. **IH. Sajal** and S. Biswas, (2023), *Bivariate Quantitative Bayesian LASSO for Detecting Association of Rare Haplotypes with Two Correlated Continuous Phenotypes*, Frontiers in Genetics, *14*.

4. **IH. Sajal**, M. Chowdhury, T. Wang, D. Euhus, P. Choudhary, and S. Biswas, (2022), *CBCRisk-Black: A Personalized Contralateral Breast Cancer Risk Prediction Model for Black Women*, Breast Cancer Research and Treatment, *194(1):179-86*.

5. R. Das Gupta, M. Akonde, **IH. Sajal**, A. Kibria, G. Muhammed, (2021), *Association between height and hypertension among US adults: analyses of National Health and Nutrition Examination Survey 2007–18*, Clinical Hypertension, *27(1), 1-12*.

6. A. Talukder, R. Das Gupta, MR. Hashan, SS. Haider, **IH. Sajal**, M. Sarker, (2021), *Association between television viewing and overweight and obesity among women of reproductive age in Timor- Leste: evidence from the demographic health survey 2016*, BMJ Open, *11(8), e045547*.

7. R. Das Gupta, M. Jahan, M. Hasan, I. Sutradhar, **IH. Sajal**, SS. Haider, M. Sarker, (2020), *Factors associated with tobacco use among Nepalese men aged 15–49 years: data from Nepal demographic and Health Survey 2016*, Clinical Epidemiology and Global Health, *8(3), 748-757*.

8. R. Das Gupta, SS. Haider, I. Sutradhar, MR. Hashan, **IH. Sajal**, M. Sarker, (2019), *Association of frequency of television watching with overweight and obesity among women of reproductive age in India: evidence from a nationally representative study*, PloS one, *14(8), e0221758*.

9. R. Das Gupta, SS. Haider, MR. Hashan, M. Hasan, I. Sutradhar, **IH. Sajal**, H. Joshi, MR. Haider, M. Sarker, (2019), *Association between the frequency of television watching and overweight and obesity among women of reproductive age in Nepal: analysis of data from Nepal Demographic Health survey 2016*, PloS one, *15(2), e0228862*.

10. R. Das Gupta, **IH. Sajal**, M. Hasan, I. Sutradhar, MR. Haider, M. Sarker, (2019), *Frequency of television viewing and association with overweight and obesity among women of the reproductive age group in Myanmar: results from a nationwide cross-sectional survey*, BMJ Open, *9(3), e024680*.

Scholarships and Awards

- **Fellows Award for Research Excellence** *National Institutes of Health* — Jul. 2025
- **Mei Lein Fellowship** *University of Texas at Dallas* — May 2023, May 2022
- **Travel Award** — Jan. 2023
  Advances in Statistical and Computational Methods
  for Analysis of Biomedical, Genetic, and Omics Data, Dallas, TX
- **Boyd Harshburger Travel Award** *Southern Regional Conference on Statistics, Jekyll Island, GA* — Sep. 2022
- **PhD Research Small Grant Program** *University of Texas at Dallas* — May 2022
- **Summer Institute in Statistical Genetics Scholarship** *University of Washington* — May 2021
- **Academic Excellence Award** *University of Dhaka, Bangladesh* — Dec. 2017

Training and Workshops attended

- **Python for Data Science Workshop** *University of Texas at Dallas* — Jul. 2022
- **Summer Institute in Statistical Genetics** *University of Washington* — Jul. 2021
  Module: Introduction to Genetics and Genomics, Genetic Epidemiology, MCMC for Genetics.

Volunteer Experiences

- **Advances in Statistical and Computational Methods**
  for Analysis of Biomedical, Genetic, and Omics Data. *Dallas, TX* — Mar. 17-19, 2023
- **Florence Nightingale Day** (Event to promote statistics to school-going students.) *Dallas, TX* — 2021-2022
- **Global Conference on Implementation Science (GCIS)** *Dhaka, Bangladesh* — Jun. 29 - Jul. 1, 2019
- **International Conference on Applied Statistics (ICAS)** *Dhaka, Bangladesh* — Dec. 27-29, 2014

Professional Affiliation

Member - **American Statistical Association** — Jun. 2021 - present

Associate Member - **American Association for Cancer Research** — Jan. 2026 - present