

# Jiacong Du

734-5466369; [jiacong@umich.edu](mailto:jiacong@umich.edu)

## Education

---

University of Michigan, Ann Arbor

- PhD in Biostatistics (GPA: 4.0/4.0) 2020-2025 summer(expected)
- Master in Biostatistics (GPA: 4.0/4.0) 2018-2020

Peking University, China

- BA in Biomedical English; BS in Mathematics and Applied Mathematics 2013-2018

## Internship

---

*PhD Statistician Intern*

May – Aug 2024

*Advisor: Dr. Bochao Jia, Dr. Jonathan Denne*

*Eli Lilly and Company*

- Developed a statistical tool for predicting and identifying patient dropout in on-going clinical trials, which is based on dynamic risk prediction models including partly conditional models and landmarking models.
- Our tool increased the probability of identifying true cases about 5 times compared to random selection and could potentially reduce 20-30% patient dropout.

## Research projects

---

*Longitudinal data analysis with informative visiting process (Thesis, on-going)*

2022-current

*Advisor: Dr. Xu Shi, Dr. Bhramar Mukherjee*

*University of Michigan*

- Conducted literature review for analyzing longitudinal data in the Electronic Health Records (EHRs). Performed comprehensive simulations to demonstrate the impact of the informative visiting process.
- Pioneered a novel method that incorporates the visiting process into joint modeling of both the longitudinal and disease processes targeting coefficient estimation.

*Data integration in causal inference with high-dimensional data (Thesis)*

2022-current

*Advisor: Dr. Xu Shi, Dr. Bhramar Mukherjee*

*University of Michigan*

- Proposed a novel method to integrate EHR data with survey samples for estimating the average treatment effect (ATE). The proposed method corrects selection bias in the EHRs and provides doubly robust inference for the ATE with high-dimensional covariates.
- Implemented a Newton-Raphson algorithm for solving the penalized estimating equation in C++ and Rcpp.
- Applied the proposed method to integrate EHRs from ~80,000 participants in the Michigan Genomics Initiative with survey data from ~4,000 participants in the NHANES, incorporating ~60 covariates.

*Large-scale single mediator hypothesis testing*

2019-2022

*Advisor: Dr. Bhramar Mukherjee, Dr. Jennifer A. Smith*

*University of Michigan*

- Developed a novel Sobel-comp method for testing mediation effects with high-dimensional mediators. The proposed method demonstrates greater statistical power compared to traditional methods.
- Developed an R package [medScan](#) on CRAN.
- Evaluated the mediation mechanism from adult socioeconomic status to HbA1c levels through DNA methylation (approximately 800,000 methylation CpG sites) using data from MESA.

*Variable selection with multiply imputed datasets*

2019-2022

*Advisor: Dr. Bhramar Mukherjee*

*University of Michigan*

- Provided a framework of regularized methods to ensure consistent variable selection across multiple imputed datasets, streamlining the variable selection process after missing value imputation.
- Derived cyclic coordinate descent and majorization-minimization optimization algorithms.
- Developed an R package [miselect](#) on CRAN.

## Honors/Awards

---

- Best presentation award at the 2024 Michigan Student Symposium for Interdisciplinary Statistical Sciences, 2024
- ENAR 2024 Spring Meeting RAB poster award (**1<sup>st</sup> place**), 2024
- Academic Excellence Award of Peking University, 2017
- Tung OOCL Scholarship of Peking University (**Top 2**), 2014

## Skills

---

- C++, R, statistical modeling, machine learning, causal inference, missing data, variable selection.

## Packages/software

---

- `miselect`: R package for variable selection with multiply-imputed datasets.
- `medScan`: R package for single-mediator hypothesis testing for a large number of mediators.
- [R shiny app](#): for exploring test allocation strategies across a variety of pandemic scenarios.

## Publications

---

- Fritsche, L. G., Nam, K., **Du, J.**, Kundu, R., Salvatore, M., Shi, X., ... & Mukherjee, B. (2023). Uncovering associations between pre-existing conditions and COVID-19 Severity: A polygenic risk score approach across three large biobanks. *PLoS genetics*, *19*(12), e1010907.
- **Du, J.**, Zhou, X., Clark-Boucher, D., Hao, W., Liu, Y., Smith, J. A., & Mukherjee, B. (2023). Methods for large-scale single mediator hypothesis testing: Possible choices and comparisons. *Genetic Epidemiology*, *47*(2), 167-184.
- Clark-Boucher, D., Zhou, X., **Du, J.**, Liu, Y., Needham, B. L., Smith, J. A., & Mukherjee, B. (2023). Methods for mediation analysis with high-dimensional DNA methylation data: Possible choices and comparisons. *PLoS genetics*, *19*(11), e1011022.
- **Du, J.**, Boss, J., Han, P., Beesley, L. J., Kleinsasser, M., Goutman, S. A., ... & Mukherjee, B. (2022). Variable selection with multiply-imputed datasets: choosing between stacked and grouped methods. *Journal of Computational and Graphical Statistics*, *31*(4), 1063-1075.
- Wang, Y. Z., Zhao, W., Ammous, F., Song, Y., **Du, J.**, Shang, L., ... & Smith, J. A. (2022). DNA methylation mediates the association between individual and neighborhood social disadvantage and cardiovascular risk factors. *Frontiers in Cardiovascular Medicine*, *9*, 848768.
- **Du, J.**, J Beesley, L., Lee, S., Zhou, X., Dempsey, W., & Mukherjee, B. (2022). Optimal diagnostic test allocation strategy during the COVID-19 pandemic and beyond. *Statistics in Medicine*, *41*(2), 310-327.
- Ray, D., Salvatore, M., Bhattacharyya, R., Wang, L., **Du, J.**, Mohammed, S., ... & Mukherjee, B. (2020). Predictions, role of interventions and effects of a historic national lockdown in India's response to the COVID-19 pandemic: data science call to arms. *Harvard data science review*, *2020*(Suppl 1).

## Preprints

---

- **Du, J.**, Shi, X., Zeng, D., & Mukherjee, B. (2024). Doubly robust causal inference through penalized bias-reduced estimation: combining non-probability samples with designed surveys. *arXiv preprint arXiv:2403.18039*. Journal of the Royal Statistical Society, Series B. Under review.
- Yu, Y., **Du, J.**, Zhang, M., Wu, Z., Ryan, A. M., & Mukherjee, B. (2022). Outcome Adaptive Propensity Score Methods for Handling Censoring and High-Dimensionality: Application to Insurance Claims. *arXiv preprint arXiv:2208.00114*. Statistical Methods and Medical Research. Accepted.