

# Jin Jin

---

## Contact Information

Department of Biostatistics, Epidemiology and Bioinformatics  
University of Pennsylvania, Perelman School of Medicine  
203 Blockley Hall  
423 Guardian Drive  
Philadelphia, PA 19104  
Phone: (612) 707 - 5890  
Email: jin.jin@pennmedicine.upenn.edu  
Personal website: <https://jin93.github.io>

## Education

**Ph.D. Biostatistics**, 2014 - 2019  
University of Minnesota, Minneapolis, MN  
Thesis: Voxel-wise detection of prostate cancer using multi-parametric MRI data.  
Advisors: *Joseph Koopmeiners*, *Lin Zhang*

**M.S. Biostatistics**, 2014 - 2016  
University of Minnesota, Minneapolis, MN

**B.S. Statistics**, 2010 - 2014  
School for the Gifted Young  
University of Science and Technology of China, Hefei, Anhui, China

## Research Interests

Bayesian statistics, statistical genetics and genomics, predictive modeling.

### Areas of applications:

Health inequity  
Statistical data integration  
Disease risk prediction and precision medicine

## Academic Appointments

**Assistant Professor of Biostatistics (tenure-track)**, Feb. 2023 - Present  
Department of Biostatistics, Epidemiology and Bioinformatics  
University of Pennsylvania

**Postdoctoral Fellow**, Sept. 2019 - Jan. 2023  
Department of Biostatistics  
Johns Hopkins Bloomberg School of Public Health  
Mentor: *Nilanjan Chatterjee*

**Research & Teaching Assistant**, 2014 - 2018  
Division of Biostatistics, School of Public Health  
University of Minnesota

## Professional Experience

**Research Scientist Intern**, May 2018 - Aug. 2018  
Department of Biostatistics and Programming  
Sanofi S.A., Cambridge, MA

**Data Science Intern**, May 2017 - Aug. 2017  
Global Specialty Analytics  
Liberty Mutual Insurance, Boston, MA

## Honors and Awards

### Awards & Scholarships

James R. Boen Student Achievement Award, University of Minnesota, 2019.  
Doctoral Dissertation Fellowship, University of Minnesota, 2018.  
COGS Grant Award, University of Minnesota, 2017.  
Dean's PhD Scholar's Award, University of Minnesota, 2016.  
John E. Connett First Year Award, University of Minnesota, 2015.  
Outstanding Student Scholarship, University of Science and Technology of China, 2012, 2013 & 2014.

### Competitions

Student Paper Award, Medical Device and Diagnostic Section, AMA Joint Statistical Meetings, 2017 & 2019.  
1<sup>st</sup> place, Ph.D. Student Award, School of Public Health Research Day, University of Minnesota, 2017.  
Silver Medal, China Girls' Mathematical Olympiad (CGMO), 2010.

## Refereed Publications

\*Authors contributed equally to the manuscript

1. Zhang, J., Zhan, J., **Jin, J.**, Ma, C., Zhao, R., Connell, J.O., Jiang, Y., 23andMe Research Team, Koelsch, B.L., Zhang, H. and Chatterjee, N. An Ensemble Penalized Regression Method for Multi-ancestry Polygenic Risk Prediction. [bioRxiv](#). To appear in *Nature Communications*.
2. **Jin, J.**, Qi, G., Yu, Z., Chatterjee, N. "Mendelian Randomization Analysis Using Multiple Biomarkers of an Underlying Common Exposure.". *Biostatistics*. Epub ahead of print. DOI: 10.1093/biostatistics/kxae006. 2024.
3. **Jin, J.**, Zhan, J., Zhang, J., Zhao, R., 23andMe Research Team, Buyske, S., Gignoux, C., Haiman, C., Kenny, E.E., Kooperberg, C., North, K., Wojcik, G., Zhang, H., Chatterjee, N. "MUSSEL: enhanced Bayesian polygenic risk prediction leveraging information across multiple ancestry groups." [bioRxiv](#). To appear in *Cell Genomics*.
4. **Jin, J.**, Chatterjee, N. "Polygenic Risk Prediction and Precision Prevention." Statistical Methods for Precision Health. Chakraborty, B., Laber, E., Moodie, E., Cai, T., Van der Laan, M.J. (Eds). Chapman & Hall/CRC. In press. 2023.

5. Zhang, H., Zhan, J., **Jin, J.**, Zhang, J., Ahearn, T., Yu, Z., O Connell, J., Jiang, Y., Koelsch, B., 23andMe research team, Lin, X., Garcia-Closas, M., Chatterjee, N. “A New Method for Multiancestry Polygenic prediction Improves Performance across Diverse Populations.” *Nature Genetics*, [Link](#), 2023.
6. Dzaye, O., Razavi, A.C., Dardari, Z.A., Wang, F., Honda, Y., Nasir, K., Coresh, J., Howard-Claudio, C. M., **Jin, J.**, Yu, B., de Vries, P., Wagenknecht, L., Folsom, A., Blankstein, B., Kelly, T.N., Whelton, S.P., Mortensen, M.B., Chatterjee, N., Matsushita, K., Blaha, M.J. “Polygenic Score and Extreme Coronary Artery Calcium Phenotypes (CAC=0 and CAC>1000) in Adults>75 Years Old: The Atherosclerosis Risk in Communities Study.” *Journal of Cardiovascular Computed Tomography*, 17(4), S66-S67, 2023.
7. Hou, X., Guo, P., Wang, P., Liu, P., Lin, D., Fan, H., Li, Y., Wei, L., Lin, Z., Jiang, D., **Jin, J.**, Pillai, J.J., Huang, J., Pinho, M.C., Thomas, B.P., Welch, B.G., Park, D.C., Hillis, A.E., Patel, V.M., and Lu, H. “Deep Learning Enables Cerebrovascular Reactivity and Hemodynamic Delay Mapping Using Resting-state fMRI.” *npj Digital Medicine*, 6(1), 116, 2023.
8. Steinbrenner, I., Yu, Z., **Jin, J.**, Schultheiss, U.T., Kotsis, F., Grams, M., Coresh, J., Wuttke, M., Eckardt, K.U., Chatterjee, N., Sekula, P., Kötgen, A., on behalf of the GCKD investigators. “A Polygenic Score for Reduced Kidney Function and Adverse Outcomes in a Chronic Kidney Disease Cohort.” *Kidney International*, 103(2), 421-424, 2023.
9. **Jin, J.**, Zhang, L., Leng, E., Metzger, G.J., Koopmeiners, J.S. “Multi-resolution Super Learner for Voxel-wise Classification of Prostate Cancer Using Multi-parametric MRI.” *J Appl Stat*, 50(3), 805-826, 2023.
10. **Jin, J.**, Yue, W. “T2-DAG: A Powerful Test for Differentially Expressed Gene Pathways via Graph-informed Structural Equation Modeling.” *Bioinformatics*, 38(4), 1005-1014, 2022.
11. Rabinowitz, J., **Jin, J.**, Kuo, S., Thrul, J., Reboussin, B., Domingue, B., Ialongo, N., Maher, B., Uhl, G. “Positive Associations between Cannabis and Alcohol Use Polygenic Risk Scores and Phenotypic Opioid Misuse among African-Americans.” *PLOS One*, 17(4), p.e0266384, 2022.
12. **Jin, J.**, Zhang, L., Leng, E., Metzger, G.J., Koopmeiners, J.S. “Bayesian Spatial Models for Voxel-wise Prostate Cancer Classification Using Multi-parametric MRI Data.” *Statistics in Medicine*, 41(3), 483-499, 2022.
13. Rabinowitz, J.A., **Jin, J.**, Kahn, G., Kuo, S.-C., Campos, A., Renteria, M., Benke, K., Wilcox, H., Ialongo, N.S., Maher, B.S., Kertes, D., Eaton, W., Uhl, G., Wagner, B.M., Cohen, D. “Genetic Propensity for Risky Behavior and Depression and Risk of Lifetime Suicide Attempt among Urban African Americans in Adolescence and Young Adulthood.” *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics Part B: Neuropsychiatric Genetics*, 186(8), 456-468, 2022.
14. Ballreich, J., **Jin, J.**, Kundu, P., Chatterjee, N. (2021). “Provider and Patient Characteristics of Medicare Beneficiaries Who Are High-Risk for COVID-19 Mortality.” *Journal of General Internal Medicine*, 36(7), 2189-2190, 2021.
15. Yu, Z., **Jin, J.**, Tin, A., Kottgen, A., Yu B., Chen J., Ballantyne, C.M., Hoogeveen, R.C., Arking, D.E., Chatterjee, N., Coresh, J., Grams, M.E., Coresh J. “Polygenic Risk Scores for Kidney Function and Their Associations with Circulating Proteome, and Incident Kidney Diseases.” *Journal of the American Society of Nephrology*, 32(12), 3161-3173, 2021. [Selected as the “Best of ASN Journals” in 2021.]

16. **Jin, J.\***, Agarwala, N.\*, Kundu, P.\*, Harvey, B., Zhang, Y., Wallace, E., Chatterjee, N. “Individual and Community-level Risk for COVID-19 Mortality in the United States.” *Nature Medicine*, **27(2)**, 264-269, 2021.
17. **Jin, J.**, Riviere, M, Luo, X., Dong, Y. “Bayesian Methods for the Analysis of Early-phase Oncology Basket Trials with Information Borrowing across Cancer Types.” *Statistics in Medicine*, **39(25)**, 3459-3475, 2020.
18. **Jin, J.**, Liu, Q., Zheng, W., Lei, Gao., Shun, Z., Lin, T.T., Dong, Y. “A Bayesian Method for Preliminary Proof of Concept in Early Phase Oncology Studies with a Basket Design.” *Statistics in Biosciences*, **12(2)**, 167-179, 2020.
19. Leng, E., Henriksen, J.C., Rizzardi, A.E., **Jin, J.**, Nam, J.W., Brassuer, B.M., Johnson, A.D., Reder, N.P., Koopmeiners, J.S., Schmechel, S.C., Metzger, G.J. “Signature Maps for Automatic Identification of Prostate Cancer from Colorimetric Analysis of H&E-and IHC-stained Histopathological Specimens.” *Scientific Reports*, **9(1)**, 1-12, 2019.
20. **Jin, J.**, Zhang, L., Leng, E., Metzger, G.J., Koopmeiners, J.S. “Detection of Prostate Cancer with Multiparametric MRI utilizing the Anatomic Structure of the Prostate.” *Statistics in Medicine*, **37**, 3214 - 3229, 2018.
21. Leng, E., Spilseth, B., **Jin, J.**, Zhang, L., Leng, E., Koopmeiners, J.S., Metzger, G.J. “Development of a Measure for Evaluating Lesionwise Performance of CAD Algorithms in the Context of MpMRI Detection of Prostate Cancer.” *Medical Physics*, **45**, 2076 - 2088, 2018.

## Manuscripts Submitted for Publication

\*Co-corresponding authors

1. Dun, Y., Chatterjee, N., **Jin, J.\***, Nishimura, A\*. A Robust Bayesian Method for Building Polygenic Risk Scores using Projected Summary Statistics and Bridge Prior. arXiv. DOI: arXiv:2401.15014. Jan 2024.
2. **Jin, J.**, Kim, E., Chatterjee, N. “Polygenic Risk Scores: Development, Applications and Implementations.” Submitted to *Clinical Chemistry*.

## Grant Support

### Currently Funded Grants

1. NIH R00-HG012223, “Multi-ethnic risk prediction for complex human diseases integrating multi-source genetic and non-genetic information”, 02/01/2022 - 01/31/2026.

**Name of PI:** Jin Jin

**Role:** Principal Investigator

**Project narrative:** The proposed research aims to improve healthcare for the populations that are underrepresented in the public health fields. The research will generate novel methodologies and user-friendly tools to improve disease risk prediction for minority populations by integrating trans-ethnic, disparate sources of information on genetic markers, clinical and environmental factors, health conditions, and socio-demographic indicators.

2. NIH U01-CA164947, “Post Genome Wide Association Studies in Testicular Germ Cell Tumors”, 09/21/2020 - 08/31/2025.

**Name of PI:** Katherine Nathanson

**Role: Co-Investigator**

**Project narrative:** Testicular germ cell tumors (TGCT) are the most common tumor type in white men ages 15-45, with an increasing incidence. They have a high heritability and inherited variation underlies 50% of the risk of the disease. We build upon our prior research in the Testicular Cancer Consortium to identify additional genetic variation associated with disease risk, so that we expand our understanding of TGCT biology and epidemiology, as well as identify the men at highest risk of disease.

**Presentations****Invited Talks**

1. “Gene Regulatory Network-Informed Identification of Differentially Expressed Gene Pathways and Disease Classification”, 20th Annual Meeting of MCBIOS, Atlanta, GA, Mar. 2024.
2. “Mendelian Randomization Analysis of Latent Exposures Co-regulating Multiple Correlated Traits”, ENAR Spring Meetings, Baltimore, MD, Mar. 2024.
3. “MUSSEL: Enhanced Bayesian Polygenic Risk Prediction Leveraging Information across Multiple Ancestry Groups” (Virtual talk), Statistical Genetics and Genomics Working Group, Columbia University, Dec. 2023.
4. “MUSSEL: Enhanced Bayesian Polygenic Risk Prediction Leveraging Information across Multiple Ancestry Groups” (Virtual talk), Statistical Genetics and Genomics Working Group, Columbia University, Dec. 2023.
5. “MUSSEL: Enhanced Bayesian Polygenic Risk Prediction Leveraging Information across Multiple Ancestry Groups” (Virtual talk), Polygenic Risk Methods in Diverse populations (PRIMED) Methods Development Sub-WG Meeting, Oct. 2023.
6. “Information Borrowing Approaches for Developing Equitable Polygenic Prediction Models across Diverse Ancestries” (Virtual talk), the 6th International Conference on Econometrics and Statistics (EcoSta 2023), Tokyo, Japan, Aug. 2023.
7. “Information Borrowing Approaches for Developing Equitable Polygenic Prediction Models across Diverse Ancestries” (Virtual talk), 2023 ISI World Statistics Congress, Ottawa, Canada, July 2023.
8. “ME-Bayes SL: Enhanced Bayesian Polygenic Risk Prediction Leveraging Information across Multiple Ancestry Groups”, ICSA 2023 China Conference, Chengdu, China, July 2023.
9. “ME-Bayes SL: Enhanced Bayesian Polygenic Risk Prediction Leveraging Information across Multiple Ancestry Groups”, 2023 WNAR/IMS annual meeting, Anchorage, Alaska, June 2023.
10. “Developing and Validating Prediction Models for Individual- and Community-level Risk for COVID-19 Mortality”, 2023 Lifetime Data Science (LiDS) conference, Raleigh, NC, June 2023.
11. “Genetic Propensity for Risky Behavior and Depression and Risk of Lifetime Suicide Attempt among Urban African Americans in Adolescence and Young Adulthood”. The 2023 Thomas R. Ten Have Symposium on Statistics in Mental Health, Boston, MA, May 2023.
12. “Disease Risk Prediction through the Integration of Information across Disparate Data Sources”, Department of Human Genetics, the University of Chicago, Chicago, IL, Feb. 2022.

13. “Disease Risk Prediction through the Integration of Information across Disparate Data Sources”, Department of Biostatistics and Bioinformatics, Milken School of Public Health, George Washington University, Virtual, Feb. 2022.
14. “Disease Risk Prediction through the Integration of Information across Disparate Data Sources”, Department of Biostatistics and Bioinformatics, School of Medicine, Duke University, Virtual, Feb. 2022.
15. “Disease Risk Prediction through the Integration of Information across Disparate Data Sources”, Department of Biostatistics, Epidemiology and Bioinformatics, Perelman School of Medicine, University of Pennsylvania, Virtual, Feb. 2022.
16. “Individual- and Community-Level Disease Risk Prediction through the Integration of Information across Disparate Data Sources”, Public Health Science Division, Fred Hutchinson Cancer Research Center, Virtual, Feb. 2022.
17. “Mendelian Randomization Analysis Using Multiple Biomarkers of an Underlying Common Exposure”, ICSA Applied Statistics Symposium, Sept. 2021.
18. “Estimating the Size of High-risk Populations for COVID-19 Mortality Across 442 US Cities”, Department of Biostatistics & Bioinformatics, School of Public Health, University of Colorado, Denver, June 2020.
19. “Voxel-wise Classification of Prostate Cancer Using Multi-parametric MRI Data”, 2020 Global Scientist Interdisciplinary Forum, Shenzhen, China, Jan. 2020.

#### **Contributed Conference Presentations**

1. “Mendelian Randomization Analysis of Latent Exposures Co-regulating Multiple Correlated Traits”. ENAR Spring Meetings, Baltimore, MD, Mar. 2024.
2. “Ancestry-specific Polygenic Risk Scores for Telomere Length in TOPMed.” Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Meeting (Virtual), Oct. 2021.
3. “Mendelian Randomization Analysis for the Causal Effect of a Latent Exposure Underlying Multiple Related Traits.” American Society of Human Genetics Annual Meeting (Virtual), Oct. 2020.
4. “Bayesian Hierarchical Models for Voxel-Wise Classification of Prostate Cancer Accounting for Spatial Correlation and Between-Patient Heterogeneity in the Multi-Parametric MRI Data.” Joint Statistical Meetings, Denver, CO, July 2019.
5. “Bayesian Hierarchical Models for Voxel-wise Prostate Cancer Classification Using Nearest Neighbor Gaussian Process.” Joint Statistical Meetings, Vancouver, BC, Canada, July 2018.
6. “Detection of Prostate Cancer with Multi-parametric MRI Models Utilizing the Anatomic Structure of the Prostate.” Joint Statistical Meetings, Baltimore, MD, Aug. 2017.
7. “Multi-Parametric MRI Models for Prostate Cancer Diagnosis under a Bayesian Hierarchical Model Framework.” ENAR Spring Meetings, Washington, DC, Mar. 2017.

## Teaching

### **Instructor** (*at University of Pennsylvania*)

BSTA 6210: Statistical Inference I, Spring 2024

### **Invited Lectures**

Polygenic Risk Prediction: Algorithms, Fairness and Applications, BSTA 787 Guest Lecture, Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Feb. 2024

Linear and Logistic Regression, FDA Pharmacoeconomics Certificate Program Biostatistics Course, Dec. 2023

Polygenic Risk Prediction: Algorithms, Fairness and Applications, BSTA 7980 Guest Lecture, Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Sept. 2023

### **Teaching Assistant** (*at the University of Minnesota*)

Bayesian Decision Theory and Data Analysis, Spring 2018

Advanced Statistical Inference, Spring 2017

Survival Analysis, Fall 2016 & Fall 2017

Biostatistics Modeling and Methods, Spring 2016

Fundamentals of Biostatistical Inference, Fall 2015

Biostatistics II, Spring 2015

Biostatistical Methods I, Fall 2014 & Spring 2015

## Advising/Mentoring

(*at the University of Pennsylvania*)

### **PhD Advising**

Dong Heon Lee, Ph.D. student in Biostatistics, 2023 - present

### **PhD Committees**

Committee Member, Katrina Bazemore, Ph.D. student in Epidemiology, 2023 - present

Committee Chair, Zhuoping Zhou, Ph.D. student in Applied Mathematics and Computational Science (AMCS), 2024 - present

### **Master's Advising**

Yihao Wang, MS student in Biostatistics, 2023 - present

### **Trainees/Research Assistants Supervised**

Andrew Lakkis (BGS lab rotation), Ph.D. student in Biostatistics, 2023 Fall

Yue Yu, MS student in Biostatistics, 2023 - present

Ruofan Wang, MS student in Biostatistics, 2023 - present

### **Academic Advising**

Chunyu Luo, MS student in Biostatistics, 2023 - present

## Professional Activities

### Memberships in Professional and Scientific Societies:

American Statistical Association (ASA), 2017 - present, Member

International Biometrics Society - Eastern North American Region (ENAR), 2017 - present, Member

International Biometrics Society - Western North American Region (WNAR), 2020 - present, Member

American Society for Human Genetics (ASHG), 2020 - present, Member

International Chinese Statistical Association (ICSA), 2020 - present, Member

Patient-Centered Outcomes Research Institute (PCORI), 2023 - present, Grant Reviewer

### Service to Professional Societies:

Organizer and Chair, Invited Session on Recent Advances in High-dimensional Statistical Inference, 2024 ICSA China Conference, Wuhan, China, June 2024.

Reviewer, ASA Section on Statistics in Genomics and Genetics' Distinguished Student Paper Competition, 2024.

Organizer and Chair, Invited Session on Integrative Analysis of Multi-Modal Data with Applications in Risk Prediction, Neuroimaging, and Genomic Studies, 2023 WNAR/IMS annual meeting, Anchorage, AK, June 2023.

Moderator, Platform Session on Considering Ancestry in Study Populations, 2022 Annual Meeting of the American Society of Human Genetics (ASHG), Los Angeles, CA, Oct. 2022.

Organizer and Chair, Invited Session on Recent Advancements In Statistical Data Integration, 2022 ICSA Applied Statistics Symposium, Gainesville, FL, June 2022.

Contributed Session Chair, Section on Medical Devices & Diagnostics, Joint Statistical Meetings, Vancouver, BC, Canada, Aug. 2018.

### Referee for:

Annals of Human Genetics

Annals of Translational Medicine

Biostatistics

Cancers

Cancer Epidemiology, Biomarkers & Prevention

Contemporary Clinical Trials

International Journal of Epidemiology

Journal of the Royal Statistical Society: Series A

Journal of Clinical Pediatrics and Neonatology

Journal of Spatial Science

Nature Communications

New England Journal of Statistics in Data Science

Pharmaceutical Statistics

PLOS Computational Biology

Statistics in Medicine

Translational Cancer Research



## **University/Departmental Service:**

*at the University of Pennsylvania*

Member, Biostatistics Qualify Exam Committee, 2024 - present

Member, Biostatistics Student Admissions Committee, 2023 - present

Member, GGEB Student Recruitment Committee, 2023 - present

Member, GGEB Awards Committee, 2023 - present

Member, Summer Undergraduate Internship Program (SUIP) committee, 2023 - present

## **Software**

### **R Packages**

#### **MRLE**

Mendelian randomization analysis for latent exposures leveraging information from multiple biomarkers.

#### **T2DAG**

A DAG-informed high-dimensional two-sample test for trait-associated gene pathways.

#### **MUSSEL**

An R-based command line tool for implementing MUSSEL, a powerful method for developing ancestry-specific polygenic risk score (PRS) that integrates information from GWAS summary statistics and external LD reference data from multiple populations (ancestry groups).

### **COVID-19 Risk Tools**

#### **Mortality Risk Calculator**

The tool provides an assessment of individualized risks for mortality from COVID-19 using the best publicly available information on risks associated with various pre-existing conditions and socio-demographic factors. We further tailored it to produce absolute risk estimates in future time frames by incorporating information on pandemic dynamics at the community level with projections available from an ensemble of pandemic forecasting models.

#### **Risk Interactive Maps**

Developed through a collaborative effort with *PolicyMap*, the interactive maps show sizes of high-risk populations in the U.S., nationwide and by cities, countries, and states.