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 modeling, statistical genetics and genomics, predictive modeling, Mendelian randomization.

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.

1st 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, 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. "Novel Methods for Multi-ancestry Polygenic Prediction and their Evaluations in 3.7 Million Individuals of Diverse Ancestry." bioRxiv. To appear in *Nature Genetics*.

2. **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.

3. 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." medRxiv. To appear in *Kidney International*.

4. Steinbrenner, I., Yu, Z., **Jin, J.**, Schultheiss, U.T., Kotsis, F., Grams, M., Coresh, J., Wuttke, M., Eckardt, K.U., Sekula, P., Chatterjee, N., Kötgen, A. "MO517: A Polygenic Risk Score for Reduced EGFR is Associated With Adverse Events in a Chronic Kidney Disease Cohort the German Chronic Kidney Disease Study." *Nephrology Dialysis Transplantation*, 37, gfac071-048. DOI: https://doi.org/10.1093/ndt/gfac071.048.

5. **Jin, J.**, Yue, W. "T2-DAG: A Powerful Test for Differentially Expressed Gene Pathways via Graph-informed Structural Equation Modeling." *Bioinformatics*, 38(4), pp.1005-1014.

6. **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*, 2021. DOI: 10.1080/02664763.2021.2017411.

7. Rabinowitz, J., **Jin, J.**, Kuo, S., Thrul, J., Reboussin, B., Domingue, B., Ialongo, N., Maher, B., Uhl, G. "Associations between Cannabis and Alcohol Consumption Polygenic Risk Scores and Ever Misusing Opioids in an Urban, African American Cohort." To appear in *PLOS One*.

8. **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*, 1-17, 2021. DOI: 10.1002/sim.9245.

9. 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*, 1-13, 2021. DOI: 10.1002/ajmg.b.32866.

10. Ballreich, J., **Jin, J.**, Kundu, P., Chatterjee, N. "Provider and Patient Characteristics of Medicare Beneficiaries Who Are High-Risk for COVID-19 Mortality." *Journal of General Internal Medicine*, **36**, 21892190, 2021. DOI: 10.1007/s11606-021-06857-8.

11. 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 to the Circulating Proteome, and Incident Kidney Diseases: the Atherosclerosis Risk in Community Study." *Journal of the American Society of Nephrology*, 2021. DOI: 10.1681/ASN.2020111599. [This work was selected by the American Society of Nephrology (ASN) as the "Best of ASN Journals" in 2021.]

12. **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, 2020.

13. **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.

14. **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.

15. 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.

16. **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.

17. 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

1. **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. "ME-Bayes SL: enhanced Bayesian polygenic risk prediction leveraging information across multiple ancestry groups." bioRxiv. Under review at *Cell Genomics*.

2. 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. Under review at *Nature Communications*.

3. Dzaye, O., Razavi, A.C., Dardari, Z.A., Wang, F., Honda, Y., Nasir, K., Coresh, J., Howard-Claudio, C. M., **Jin, J.**, Boerwinkle, E., Yu, B., 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." Under review at Circulation. Submitted to *Nature Communications*.

4. 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." arXiv. Submitted to *npj Digital Medicine*.

5. **Jin, J.**, Qi, G., Yu, Z., Chatterjee, N. "Mendelian Randomization Analysis Using Multiple Biomarkers of an Underlying Common Exposure." bioRxiv. Under review at *Biostatistics*.

Manuscripts in Preparation

1. Fu, S., Dutta, D., Li, X., Ahearn, T.U., Li, Z., Wang, K., **Jin, J.**, Li, X., Michailidou, K., Milne, R.L., Kraft, P.,Simard, J., Pharoah, P.D.P., Schmidt, M.K., Easton, D.F., Chatterjee, N., Garcia-Closas, M., Yu, K., Zhang, H. on behalf of the Breast Cancer Association Consortium. "Accounting for cancer subtypes heterogeneity identifies eight breast cancer risk loci."

2. **Jin, J.**, Kundu, P., Chatterjee, N. "A Quasi-likelihood-based Bayesian Framework for the Integration of Multiple Regression Models across Studies with Disparate Covariate Information."

3. **Jin, J.**, Taub, M., Conomos, M., Chatterjee, N., Mathias, R. "Ancestry-specific Polygenic Risk Scores for Telomere Length and a Phenome-wide Association Study for Their Association with Risks of Age-related Diseases." Slides

4. **Jin, J.**, Luo, X. "Regression-assisted Basket Trial Design."

5. Dun, Y.*, **Jin, J.***, Saha, A., Nishimura, A., Chatterjee, N. Polygenic risk prediction via a Bayesian Bridge prior.

Grant Support

Currently Funded Grant

1. NIH R00-HG012223, "Multi-ethnic risk prediction for complex human diseases integrating multi-source genetic and non-genetic information". Role: **Principal Investigator**. 100% salary support, 02/01/2022 - 01/31/2023.

   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 the minority populations by integrating trans-ethnic, disparate sources of information on the genetic markers, clinical and environmental factors, health conditions, and socio-demographic indicators.

Presentations

Invited Talks

1. "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.

2. "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.

3. "Disease Risk Prediction through the Integration of Information across Disparate Data Sources", Department of Human Genetics, the University of Chicago, Chicago, IL, Feb. 2022.

4. "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.

5. "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.

6. "Disease Risk Prediction through the Integration of Information across Disparate Data Sources", Department of Biostastistics, Epidemiology and Bioinformatics, Perelman School of Medicine, University of Pennsylvania, Virtual, Feb. 2022.

7. "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.

8. "Mendelian Randomization Analysis Using Multiple Biomarkers of an Underlying Common Exposure", ICSA Applied Statistics Symposium, Sept. 2021.

9. "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.

10. "Voxel-wise Classification of Prostate Cancer Using Multi-parametric MRI Data", 2020 Global Scientist Interdisciplinary Forum, Shenzhen, China, Jan. 2020.

Contributed Conference Presentations

1. "Ancestry-specific Polygenic Risk Scores for Telomere Length in TOPMed." Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Meeting (Virtual), Oct. 2021.

2. "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.

3. "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.

4. "Bayesian Hierarchical Models for Voxel-wise Prostate Cancer Classification Using Nearest Neighbor Gaussian Process." Joint Statistical Meetings, Vancouver, BC, Canada, July 2018.

5. "Detection of Prostate Cancer with Multi-parametric MRI Models Utilizing the Anatomic Structure of the Prostate." Joint Statistical Meetings, Baltimore, MD, Aug. 2017.

6. "Multi-Parametric MRI Models for Prostate Cancer Diagnosis under a Bayesian Hierarchical Model Framework." ENAR Spring Meetings, Washington, DC, Mar. 2017.

Teaching Experience

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

Mentoring Experience
co-advised with Dr. Nilanjan Chatterjee at Johns Hopkins University

Yuzheng Dun, Ph.D., Department of Biostatistics, JHU, Aug. 2021 - Present.
– Bayesian algorithms with accelerated Gibbs sampling for enhanced multi-ethnic polygenic risk prediction.

Wenxuan Lu, B.S., Department of Applied Mathematics and Statistics, JHU, June 2022 - Present.
– Implement state-of-art methods for constructing polygenic risk scores, set up pipelines for PolyPred+ and XPASS.
– Help with the evaluation of a novel multi-ancestry polygenic risk prediction model, CT-SLEB.

Yuqi Zhang, B.S., Department of Biomedical Engineering, JHU, June 2020 - Present.
– Projection models for the risk of mortality due to COVID-19.
– Polygenic risk prediction of posttraumatic Stress Disorder (PTSD) related traits.

Professional Activities

Memberships
American Statistical Association (ASA)

International Biometrics Society - Eastern North American Region (ENAR)

American Society for Human Genetics (ASHG)

International Chinese Statistical Association (ICSA)

Services
Invited Session Organizer and Chair, Recent Advancements In Statistical Data Integration, 2022 ICSA Applied Statistics Symposium, 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, Cancer Epidemiology, Biomarkers & Prevention, Contemporary Clinical Trials, International Journal of Epidemiology, Journal of Clinical Pediatrics and Neonatology, Journal of Spatial Science, Statistics in Medicine, Translational Cancer Research.

Programming Skills
Programming Languages: R, MATLAB, Python.

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.

MEBayesSL

An R-based command line tool for implementing ME-Bayes SL, 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.