Haoyu Zhang, Ph.D.

Department of Biostatistics, Harvard T.H. Chan School of Public Health
677 Huntington Ave, Boston, MA 02115, USA

✉ haoyuzhang@hsph.harvard.edu • 🌐 andrewhaoyu.github.io

Education

Johns Hopkins University, Department of Biostatistics   Baltimore, MD
Ph.D. in Biostatistics, *Advisors*: Nilanjan Chatterjee and Ni Zhao   Sep 2014 - May 2019
Thesis title: *Testing for Genetic Association and Building Risk Prediction Models for Cancers Incorporating Tumor Characteristics*

Zhejiang University, Department of Mathematics   Hangzhou, China
Bachelor of Science in Statistics   Sep 2010 - Jun 2014

Professional Experience

Harvard University, Department of Biostatistics   Boston, MA
Postdoctoral fellow, *Advisor*: Xihong Lin   Aug 2019 - Present
o Develop robust Mendelian Randomization methods to identify causal risk factors for complex traits
o Build models to quantify COVID-19 transmission patterns using U.S. states and county data
o Co-advise a Ph.D. student with Dr. Lin on developing new polygenic risk score (PRS) methods

National Cancer Institute   Rockville, MD
Special volunteer, *Advisors*: Nilanjan Chatterjee and Montserrat García-Closas   Aug 2019 - Present
o Develop transethnic PRS methods using empirical Bayes and super learning algorithm
o Construct PRS for the risk of breast cancer in African Americans

National Cancer Institute   Rockville, MD
Pre-doctoral fellow, *Advisor*: Montserrat García-Closas   July 2017 - May 2019
o Led analyses through multidisciplinary collaborations in the Breast Cancer Association Consortium
o Discovered 32 novel breast cancer susceptibility loci through overall and subtype-specific GWAS
o Revealed the complex etiological heterogeneity patterns of 173 known breast cancer risk loci

Johns Hopkins University, Department of Biostatistics   Baltimore, MD
Research assistant, *Advisor*: Nilanjan Chatterjee   Oct 2015 - May 2019
o Developed methods for testing genetic associations with cancer risk incorporating tumor characteristics
o Built PRS models for breast cancer subtypes incorporating tumor characteristics
o Collaborated on analyzing an Indian gallbladder cancer GWAS dataset

Johns Hopkins University, Department of Biostatistics   Baltimore, MD
Research assistant, *Advisor*: Thomas A. Louis   July 2015 - Dec 2017
o Developed nonparametric maximum likelihood model to estimate a population-averaged value

Yale University, Department of Biostatistics   New Haven, CT
Research assistant, *Advisors*: Hongyu Zhao and Can Yang   Jul 2013 - Sep 2013
o Built prediction models for environmental chemical exposure using GWAS datasets

The University of Queensland, Queensland Brain Institute   Brisbane, Australia
Research assistant, *Advisors*: Naomi Wray and Guobo Chen   Jul 2013 - Sep 2013
- Analyzed a GWAS dataset for lithium response in bipolar disorder patients

GRANTS AND FELLOWSHIPS

1. **K99 CA256513 NIH/NCI Pathway to Independence Award**, Methods for Mendelian randomization and mediation analysis using integrative genetic and genomic data for breast cancer, Role: Principal Investigator, Amount: $192,702, Jan 2021- Dec 2022

2. **Pre-doctoral fellowship**, National Cancer Institute, Rockville, MD, USA, July 2017- May 2019

HONORS AND AWARDS

Harvard University

1. American Society of Human Genetics Annual Conference (ASHG) Reviewers' Choice poster award, virtual meeting, Oct 2021

Johns Hopkins University

1. American Statistical Association Section on Statistics in Genomics and Genetics **Distinguished Student Paper Award**, Denver, CO, USA, Aug 2019

2. Delta Omega Scholarship, Baltimore, MD, USA, May 2019

3. Delta Omega Public Health Honor Society, Alpha Chapter, Baltimore, MD, USA, April 2019

4. Louis I. and Thomas D. Dublin Award, Baltimore, MD, USA, March 2019

5. DCEG Fellows Award for Research Excellence (DFARE), National Cancer Institute, Rockville, MD, USA, Nov 2018 (**only graduate student winner in 2018**)

6. Joseph Zeger Travel award to ENAR, Baltimore, MD, USA, March 2018

7. Symposium on Advances in Genomics, Epidemiology, and Statistics (SAGES) Poster Award, University of Pennsylvania, Philadelphia, PA, USA, June 2017

Zhejiang University

1. Meritorious Winner in the American Mathematical Contest in Modeling, virtual competition, 2013

2. First Class Scholarship for Outstanding Merit, Hangzhou, China, 2011

3. **National Scholarship**, Ministry of Education, Hangzhou, China, 2011 (**Top 2** %)

4. Excellent Student Honor, Hangzhou, China, 2011

5. Scholarship for Distinguished Social Work, Hangzhou, China, 2011

PUBLICATIONS

∗ represents joint first author

1. Mhatre, S., Richmond, R. C., Chatterjee, N., Rajaraman, P., Wang, Z., **Zhang, H.**, ... , Dikshit, R. P. (2021). The role of gallstones in gallbladder cancer in India: a Mendelian randomization study. Cancer Epidemiology and Prevention Biomarkers, 30(2), 396-403.

2. Zhang, Y. D., Hurson, A. N., **Zhang, H.**, ..., Chanock, S. J., Chatterjee, N., Garcia-Closas, M. (2020). Assessment of polygenic architecture and risk prediction based on common variants across fourteen cancers. *Nature communications*, 11(1), 1-13.

3. **Zhang, H.**∗, Ahearn, T. U.∗, ..., Easton, D., Chatterjee, N., Garcia-Closas, M. (2020). Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nature Genetics*, 52(6), 572-581.

4. **Zhang, H.**, Zhao, N., Ahearn, T.U., Garcia-Closas, M., Chatterjee, N. (2020). A mixed-model approach for powerful testing of genetic associations with cancer risk incorporating tumor characteristics. *Biostatistics*.

5. Koutros, S., Kogevinas, M., Friesen, M. C., ..., **Zhang, H.**, ..., Tardon, A. (2020). Diesel exhaust and bladder cancer risk by pathologic stage and grade subtypes. *Environment International*, 135, 105346.

6. Zhao, N., **Zhang, H.**, Clark, J., Maity, A., Wu, M. (2018). Composite kernel machine regression based on likelihood ratio test for joint testing of genetic and gene–environment interaction effect. *Biometrics*, 75(2), 625-637.

7. Hu, Q., Ding, Z., **Zhang, H.**, He, Y. (2018). The superior glenohumeral joint capsule alone does not prevent superior translation of the humeral head: an in vitro biomechanical study. *Arthroscopy: The Journal of Arthroscopic & Related Surgery*, 34(11), 2962-2970.

8. Derkach, A., **Zhang, H.**, Chatterjee, N. (2017). Power Analysis for Genetic Association Test (PAGEANT) provides insights to challenges for rare variant association studies. *Bioinformatics*, 34(9), 1506-1513.

9. Mhatre, S., Wang, Z., Nagrani, R., ..., **Zhang, H.**, ..., Dikshit, R., Chatterjee, N., Rajaraman, P. (2017). Common genetic variation and risk of gallbladder cancer in India: a case-control genome-wide association study. *The Lancet Oncology*, 18(4), 535-544.

PREPRINTS

∗ represents joint first author

1. Zhu, J., **Zhang, H.**, Niu, Z., Song, J., Jiang, X., 2020. Do mental health disorders influence the risk of female cancer? Results from a genetic study. *(Submitted)*

2. Ahearn, T.U.∗, **Zhang, H.**∗, ..., Garcia-Closas, M. Chatterjee, N., 2020, Common breast cancer risk loci

predispose to distinct tumor subtypes. *(Minor revision at Breast Cancer Research)*

3. Shi, A.*, Gaynor, S.M.*, Quick, C., **Zhang, H.**, Lin, X., 2021, COVID-19 Spread Mapper: A multi-resolution,unified framework and open-source tool. *(Major revision at Bioinformatics)*

4. Jiang, Z., **Zhang H.**, Ahearn T.U. , Garcia-Closas, M., Chatterjee N., Zhan, X., Zhao, N., 2021, The sequence kernel association test for multi-categoricaloutcomes with the application to a breast cancergenome-wide association study. *(Submitted)*

Presentations

Platform Talk

1. Genome-wide association study identifies 19 novel breast cancer loci from analyses accounting for subtype heterogeneity, ***ASHG Platform Talk***, Orlando, FL, USA, October 2017

Invited Talks

1. Developing transethnic polygenic risk scores through empirical Bayes and super learning algorithm, *International Chinese Statistical Association (ICSA) Applied Statistics Symposium*, virtual meeting, Sep 2021

2. Quantifying the transmission dynamics of COVID-19 using U.S. states and county level data, *Joint Statistical Meeting (JSM)*, virtual meeting, Aug 2021

3. Developing transethnic polygenic risk scores through empirical Bayes and super learning algorithm, *Program in Genetic Epidemiology and Statistical Genetics (PGSG) seminar series*, virtual meeting, May 2021

4. Quantifying the transmission dynamics of COVID-19 using U.S. states and county level data, *ICSA Applied Statistics Symposium*, virtual meeting, Dec 2020

5. Methods for developing ancestry specific polygenic risk scores, *Breast Cancer Association Consortium (BCAC) 25th annual meeting* , virtual meeting, Sep 2020

6. A mixed-model approach for powerful testing of genetic associations with cancer risk incorporating tumor characteristics, *Shanghai Jiao Tong University*, Shanghai, China, December 2019

7. Subtypes specific GWAS and risk prediction modeling on breast cancer subtypes, *BCAC 23rd annual meeting* , Springdale, UT, USA, April 2019

8. Standard approach and methods accounting for subtype heterogeneity identified 32 novel breast cancer loci, *BCAC 22nd annual meeting*, Edinburgh, UK, June 2018

9. Testing for genetic associations in case-control studies incorporating multivariate disease characteristics, *National Cancer Institute, Division of Cancer Epidemiology & Genetics, Integrative Tumor Epidemiology Branch*, Rockville, MD, USA, September 2017

Contributed Talks

1. Robust Mendelian randomization (MR) analysis with weak instruments, *JSM*, virtual meeting, Aug 2020

2. A mixed-model approach for powerful testing of genetic associations with cancer risk incorporating tumor characteristics, *JSM*, Denver, CO, USA, Aug 2019

3. Developing polygenic risk prediction models for cancer subtypes incorporating multivariate disease characteristics, *JSM*, Vancouver, BC, Canada, July 2018

4. The mixture approach to estimating a weighted population-averaged Value, *Eastern North American Regional meeting of the International Biometric Society (ENAR)*, Atlanta, GA, USA, March 2018

5. The mixture approach to estimating a population-averaged Value, *ENAR*, Washington, D.C., USA, March 2017

6. Testing for genetic associations in case-control studies incorporating multivariate disease characteristics, JSM, Baltimore, MD, USA, August 2017

Posters

1. Developing transethnic polygenic risk scores through empirical Bayes and super learning algorithm, *ASHG*, virtual meeting, Oct 2021

2. A mixed-model approach for powerful testing of genetic associations with cancer risk incorporating tumor characteristics, *ASHG*, San Diego, CA, USA, October 2018

3. Genome-wide association study identifies 19 novel breast cancer loci from analyses accounting for subtype heterogeneity, *Johns Hopkins Biostatistics Retreat*, Philadelphia, PT, USA, April 2018

4. Genome-wide association study identifies 19 novel breast cancer loci from analyses accounting for subtype heterogeneity, *14TH Annual NIH Graduate Student Research Symposium*, Bethesda, MD, USA, February 2018

5. Genome-wide association study identifies 19 novel breast cancer loci from analyses accounting for subtype heterogeneity, *Delta Omega Poster Competition*, Baltimore, MD, USA, Feb 2018

6. Genome-wide association study identifies 19 novel breast cancer loci from analyses accounting for subtype heterogeneity, *The Maryland Genetics, Epidemiology, and Medicine Training Program (MD-GEM)*, Baltimore, MD, USA, February 2018

7. Methods for analyzing genome-wide association Studies, *Johns Hopkins University, Department of Biostatistics, 100 years celebration*, Baltimore, MD, USA, July 2017

8. Testing for genetic association in case-control studies incorporating multivariate disease characteristics, *SAGES*, Philadelphia, PT, USA, June 2017

9. Testing for genetic association in case-control studies incorporating multivariate disease characteristics, *MD-GEM*, Baltimore, MD, USA, February 2017

Advising and Teaching

Advising

1. Tony Chen, Ph.D. student, Department of Biostatistics, Harvard University, April 2021 - Present (co-advise with Xihong Lin)
   o Developing PRS methods to improve risk predictions for complex traits
   o Performing mediation analysis using PRS as mediators

Lab Instructor

Responsibilities: teaching 3 to 4 hours labs per week, beta-test, proctor and grade exams

1. Statistical Methods in Public Health II-III (Biostat 622-623), Master Level Class, taught by Marie Diener-West and Karen Bandeen-Roche, Baltimore, MD, USA, Oct 2018 - Mar 2019

2. Statistical Methods in Public Health II-III (Biostat 622-623), Master Level Class, taught by Marie Diener-West and Karen Bandeen-Roche, Baltimore, MD, USA, Oct 2017 - Mar 2018

Teaching Assistant

Responsibilities: hold office hour, grade homework and exams, and hold some classes

1. Statistical Methods in Public Health I-IV (Biostat 621-624), Master Level Class, taught by Marie Diener-West, Karen Bandeen-Roche, John McGready and James Tonascia , Baltimore, MD, USA, Sep 2016 - May 2017

2. Statistical Theory I-IV (Biostat 731-734), PhD Level Class, taught by Elizabeth L. Ogburn, Constantine Frangakis and Michael Rosenblum, Baltimore, MD, USA, Sep 2015 - May 2016

3. Survey Methodology, Undergraduate Level Class, taught by Peng Zhang, Hangzhou, China, Feb 2014 - Jun 2014

Software and Programming Skills

R shiny app

1. **PAGEANT**: Power Analysis for GEnetic AssociatioN Tests
   Software link: https://andrewhaoyu.shinyapps.io/PAGEANT/
   Github link: https://github.com/andrewhaoyu/PAGEANT

R package

1. **CKLRT**: Composite Kernel Machine Regression Based on Likelihood Ratio Test
   CRAN link: https://cran.r-project.org/web/packages/CKLRT/index.html
   Github link: https://github.com/andrewhaoyu/CKLRT

2. **TOP**: Two-stage polytomous logistic regression

Github link: https://github.com/andrewhaoyu/TOP

Programming skills

R (80k-90k lines), C (10k lines), Shell Script, RcppEigen, Stan, Photoshop, MATLAB, PLINK, QCTOOL, GTOOL, STATA, SPSS, E-View, Origin

Editorial Activities

Ad hoc reviewer
Nature Communication, Plos Computational Biology, Plos One

Services

1. Statistical genetics group website maintainer (http://jhustatgen.org), Johns Hopkins University, Department of Biostatistics, Baltimore, MD, USA, August 2018 - May 2019

2. Session chair, Statistical Advances in the Design and Analysis of Sequence-Based Genetic Association Studies, JSM, Baltimore, MD, USA, August 2018

3. **Student representative**, Johns Hopkins University, Department of Biostatistics, Baltimore, MD, USA, August 2017 - August 2018
   o Hold the student lunches with topics including passing qualify exam and finding summer internship
   o Organizes sessions to help students writing cover letter for summer internship
   o Attend faculty monthly meetings

4. TA training day planning committee, Johns Hopkins University, Department of Biostatistics, Baltimore, MD, USA, August 2018

5. **Johns Hopkins Biostatistics Mixer Organizer**, JSM, Vancouver, BC, Canada, August 2018

6. **Johns Hopkins Biostatistics Mixer Organizer**, ENAR, Atlanta, GA, USA, August 2018

7. TA training day planning committee, Johns Hopkins University, Department of Biostatistics, Baltimore, MD, USA, August 2017

8. TA training day planning committee, Johns Hopkins University, Department of Biostatistics, Baltimore, MD, USA, August 2016

9. **Volunteer service**, Sharadadevi Gramudyog Utpadak Mandali Ltd (SGS), **Baroda, India**, June 2011 - August 2011
   o Collaborated with volunteers form 12 countries to create a detailed survey to examine multiple aspects of village life
   o Interviewed locals in 16 villages to gather data about education, hygiene, income and saving, etc.
   o Analyzed collected data and presented recommendations about serving the needs of nearby communities