Ryan Sun

CONTACT INFORMATION

Department of Biostatistics
University of Texas M.D. Anderson Cancer Center
1400 Pressler Street, Unit 1410
Houston, Texas 77030

Office: (713) 745-0373
E-mail: rsun3@mdanderson.org
Website: https://ryanrsun.github.io/
Citizenship: United States, Canada

EDUCATION

Harvard University, Boston, Massachusetts

Ph.D., Biostatistics, May 2017

- Dissertation Title: "Methods for High-Dimensional Inference in Genetic Association Studies"
- Advisor: Professor Xihong Lin

M.A., Biostatistics, May 2014

Columbia University, New York City, New York

B.S., Applied Mathematics, May 2012

PROFESSIONAL EXPERIENCE

Empiric Asset Management, New York, New York

Analyst   **May 2011 - May 2012**
Quantitative analyst for hedge fund with equity market neutral strategy.

HONORS AND AWARDS

Travel Award for IMS New Researchers Conference, 2018

Harvard Program in Quantitative Genomics Postdoctoral Travel Award, 2017

ENAR Distinguished Student Paper Award, 2017

Harvard University Distinction in Teaching Award, 2014-2015

Cyprus National Government Environmental Health Travel Scholarship, 2013

Columbia University Lauren P. Breakiron Scholarship, 2008-2012

ACADEMIC EXPERIENCE

Rice University, Houston, Texas

Adjunct Assistant Professor, Department of Statistics   **July 2020 - Present**

University of Texas MD Anderson Cancer Center, Houston, Texas

Assistant Professor, Department of Biostatistics   **July 2019 - Present**

Harvard School of Public Health, Boston, Massachusetts

Postdoctoral Research Fellow   **May 2017 - June 2019**
Developed statistical methods for the analysis of high-dimensional biomedical data. Collaborations included the EPA Superfund project and the International Lung Cancer Consortium (ILCCO).

Harvard School of Public Health, Boston, Massachusetts

Instructor   **July 2016 - August 2016**
Redesigned and delivered real analysis course for incoming Biostatistics Ph.D. students. Planned lectures, created class materials, and taught all sessions.

Teaching Assistant   **September 2012 - June 2016**
Assigned to three doctoral-level biostatistics courses and one introductory statistics course.

PAPERS RECEIVING PEER-REVIEW

1. **Sun, R.***, Bouchard, M.B.*, Hillman, E.M.C. SPLASSH: Open source software for camera-based high-speed, multispectral in-vivo optical image acquisition. *Biomedical Optics Express* 2010; 1(2): 385-397.

2. Wang, Z., Claus-Henn, B., Wang, C., Wei, Y., Su, L., **Sun, R.,** Chen, H., Wagner, P.J., Lu, Q., Lin, X., Wright, R., Bellinger, D., Kile, M., Mazumdar, M., Tellez-Rojo, M.M., Schnaas, L., Christiani, D.C. Genome-wide gene by Pb exposure interaction analysis identifies UNC5D as a candidate gene for neurodevelopment. *Environmental Health* 2017; 16(1): 81.

3. **Sun, R.,** Carroll, R.J., Christiani, D.C., Lin, X. Testing for gene-environment interaction under exposure misspecification. *Biometrics* 2018; 74(2): 653-662.

4. Orkaby, A., Rich, M.W., **Sun, R.,** Lux, E., Wei, L.J., Kim, D.H. Pravastatin for primary prevention in older adults: restricted mean survival time analysis. *Journal of the American Geriatrics Society* 2018; 66(10): 1987-1991.

5. **Sun, R.,** Wang, Z., Claus Henn, B., Su, L., Lu, Q., Lin, X., Wright, R., Bellinger, D., Kile, M., Mazumdar, M., Tellez-Rojo, M.M., Schnaas, L., Christiani, D.C. Identification of novel loci associated with infant cognitive ability. *Molecular Psychiatry* 2018; 1-10.

6. **Sun, R.,** Hui, S., Bader, G., Lin, X., Kraft, P. Powerful gene set analysis in GWAS with the Generalized Berk-Jones statistic. *PLOS Genetics* 2019; 15(3): e1007530.

7. Gaynor S.*, **Sun, R.*,** Lin, X., Quackenbush, J. Identification of differentially expressed gene sets using the Generalized Berk-Jones statistic. *Bioinformatics* 2019; 35(22): 4568-4576.

8. **Sun, R.** Lin, X. Set-based tests for genetic association using the Generalized Berk-Jones statistic. *Journal of the American Statistical Association* 2020; 115(53): 1079-1091.

9. Park, H-R., Panganibana, R.A., **Sun, R.,** Shumyatcher, M., Himes, B.E., Christiani, D.C., Lu, Q. MicroRNA-124 protects neural cells against arsenic-induced endoplasmic reticulum stress and cytotoxicity in vitro and is associated with neurodevelopmental outcomes in children. *Scientific Reports* 2020; 10(1): 1-11.

10. Li, X., Li, Z., Zhou, H., Gaynor, S., Liu, Y., Chen, H., **Sun, R.**, ..., Lin, X. Dynamic incorporation of multiple in-silico functional annotations empowers rare variant association analysis of large whole genome sequencing studies at scale. *Nature Genetics* 2020+ (in press).

11. Zhu, L., Cai, D., Li, Y., Tong, X., **Sun, R.**, Srivastava, D., Hudson, M. Maximum likelihood estimation for the proportional odds model with mixed interval-censored failure time data. *Journal of Applied Statistics* 2020+ (in press).

12. **Sun, R.,** Xu, M., Li, X., Gaynor, S., Zhou, H., Bossé, Y., Lam, S., Tsao, M., Tardon, A., Chen, C., Doherty, J., Goodman, G., Egil Bojesen, S., Teresa, M.T., Johansson, M., Field, J.K., Bickeböller, H, Wichmann, H., Risch, A., Rennert, G., Arnold, S., Wu, X., Melander, O., Brunnström, H., Marchand, L.L., Zong, X., Liu, G., Andrew, A., Duell, E., Kiemeney, L.A., Shen, H., Haugen, A., Johansson, M., Grankvist, K., Caporaso, N., Woll, P., Teare, M.D., Scelo, G., Hong, Y., Yuan, J., Lazarus, P., Schabath, M.B., Aldrich, M.C., Albanes, D., Brennan, P., Barbie, D., Mak, R., Hung, R.J., Amos, C.I., Christiani, D.C., Lin, X. Identification of inflammation and immune-related risk variants associated with squamous cell lung cancer. *Genetic Epidemiology* 2020+ (in press).

13. Deng, Q., Han, G., Puebla-Osorio, N., Ma, M.C.J., Strati, P., Chasen, B., Dai, E., Dang, M., Jain, N., Yang, H., Wang, Y., Zhang, S., Wang, R., Chen, R., Showell, H., Ghosh, S., Patchva, S., Zhang, Q., **Sun, R.**, ..., Green, M.R. Characteristics of anti-CD19 CAR T-cell infusion products

associated with efficacy and toxicity in patients with large B-cell lymphomas. *Nature Medicine* 2020+ (in press).

14. Colina, A., **Sun, R.**, Wang, H., Katz, M., Hwang, H., Lee, J., Tzeng, C., Wolff, R., Raghav, K., Overman, M. Green, M.R. Natural history and prognostic factors for localized small bowel adenocarcinoma. *ESMO Open* 2020+ (in press).

15. Kim, D., Li, X., Bian, S., Wei, L.J., **Sun, R.** Effect of dementia drugs on nursing home placement: what is the meaningful metric? *JAMA Network Open* (revision invited).

16. Alhalabi, O., Soomro, Z., **Sun, R.,** Hasanov, E., Albittar, A., Tripathy, D., Valero, V., Ibrahim, N. Outcomes of changing systemic therapy in patients with relapsed breast cancer and 1 to 3 brain metastases. *ESMO Open* (revision invited).

17. Strati, S., Ahmed, S., Furqan, F., ..., **Sun, R.,** Claussen, C., Hawkins, M., Johnson, N., Singh, P., Mistry, H., Johncy, S., Adkins, S., Kebriaei, P., Shpall, E., Green, M., Flowers, C., Westin, J., Neelapu, S. Prognostic impact of corticosteroids on efficacy of chimeric antigen receptor t-cell therapy in large b-cell lymphoma. *Blood* (revision invited).

PUBLISHED CORRESPONDENCE

1. **Sun, R.,** Horiguchi, M., and Wei, L.J. Interpreting the benefit of trifluridine/tipiracil in metastatic colorectal cancer with respect to progression-free survival and overall survival. *Journal of Clinical Oncology* 2018; 36(13): 1378.

2. **Sun, R.**, Rich, M.W., and Wei, L.J. Pembrolizumab plus chemotherapy in lung cancer. *New England Journal of Medicine* 2018; 379(11): e18.

3. **Sun, R.**, Nie, L., Huang, B., Kim, D.H., and Wei, L.J. Quantifying immunoscore performance. *The Lancet* 2018; 392(10158): 1624.

4. **Sun, R.** and Wei, L.J. Regional Hyperthermia With Neoadjuvant Chemotherapy for Treatment of Soft Tissue Sarcoma. *JAMA Oncology* 2019; 5(1): 112-113.

5. **Sun, R.,** Zhu, H., and Wei, L.J. Assessing the prognostic value of automated bone scan index in prostate cancer. *JAMA Oncology* 2019; 5(2): 270.

6. **Sun, R.**, Orkaby, A.O., Kim, D.H., Zhu, H., and Wei, L.J. Interpreting non-inferiority of biodegradable-polymer stents to durable-polymer stents. *The Lancet* 2019; 393(10184): 1932-1933.

7. **Sun, R.,** Lee, H., and Wei, L.J. Interpreting the long-term prognostic value of total mesorectal excision plane quality in rectal adenocarcinoma. *JAMA Surgery* 2019; 154(1):96.

8. **Sun, R.,** Kim, D.H., and Wei, L.J. Analysis of overall survival benefit of abemaciclib plus fulvestrant in hormone receptor-positive, ERBB-2 negative breast cancer. *JAMA Oncology* 2020; 6(7):1121-1122.

9. **Sun, R.,** Messick, C., and Wei, L.J. Quantifying benefit from two-stage turnbull-cutait pull-through coloanal anastomosis for low rectal cancer. *JAMA Surgery* 2020+ (in press).

WORKING PAPERS

1. **Sun, R.*,** Claggett, B.L.*, Tian, L., Solomon, S.D., Szczech, L., Pfeffer, M.A., and Wei, L.J. On-treatment analysis for assessing the between-group difference in a comparative randomized clinical study.

2. Han, G., Yang, G., Hao, D., Lu, Y., Thein, K., Chen, J., **Sun, R.,** ..., Wang, L. 9p21 loss confers a cold tumor immune microenvironment and primary resistance to immune checkpoint therapy.

3. Xu, M., Tapia, C., Hajjar, J., ..., **Sun, R.,** Karp, D., Koay, E., Yang, Y., Wistuba, I., Hwu, P., Meric-Bernstam, F., Naing, A. Implementation of a novel web-based lesion selection tool to improve acquisition of biopsies.

4. Byun, J., Han, Y., Li, Y., Xia, J., Xiao, X., **Sun, R.,** ..., Amos, C. Trans-ethnic genome-wide meta-analysis of 35732 cases and 34424 controls identifies novel genomic cross-ancestry loci contributing to lung cancer susceptibility.

5. Yam, C., Rauch, G., Rahman, T., Karuturi, M., Ravenberg, E., White, J., Clayborn, A., McCarthy, P., Abouharb, S., Lim, B., Litton, J., Ramirez, D., Saleem, S., Stec, J., Symmans, W., Huo, L., Damodaran, S., **Sun, R.,** Moulder, S. A phase II study of mirvetuximab soravtansine in triple-negative breast cancer.

6. **Sun, R.*,** McCaw, Z.*, Tian, L., Uno, H., Kim, D., Wei, L.J. A simple solution to problems with conventional stratified analysis for assessing the overall treatment effect in a comparative study.

Sun, R., Li, Y., Liang, Z. Variance components tests for set-based genetic inference with interval-censored outcomes.

Sun, R. and Lin, X. A unified framework for inference in mediation, pleiotropy, and replicability analyses related to genetic association studies.

Sun, R. and Lin, X. Optimality of set-based strategies in multiple outcome and multiple genotype genetic association settings.

*Indicates equal contribution as first authors.

Numbered working papers are currently submitted or have been submitted, and preprints are available upon request.

| INVITED TALKS | Set-based inference for analysis of genetic compendiums (given online due to COVID19). Human Genetics Center Seminar, UTHealth School of Public Health. Houston, Texas. October 2020. |

Set-based inference for analysis of genetic compendiums (given online due to COVID19). Statistics Department Seminar, University of Haifa. Haifa, Israel. July 2020.

Set-based inference for analysis of genetic compendiums (cancelled due to COVID19). International Chinese Statistical Association China Conference. Wuhan, China. July 2020.

Set-based inference for analysis of genetic compendiums (cancelled due to COVID19). Georgia State University 8th Workshop on Biostatistics and Bioinformatics. Atlanta, Georgia. May 2020.

Identification of inflammation and immune-related risk variants associated with squamous cell lung cancer. International Lung Cancer Consortium Annual Meeting. Toronto, Ontario, Canada. September 2018.

Set-based tests using the Generalized Berk-Jones statistic in genetic association studies. Jilin University School of Mathematics Statistics Colloquium. Changchun, Jilin, China. May 2018.

A unified framework for composite null inference in mediation, pleiotropy, and replicability analyses related to genetic association studies. Harvard School of Public Health Program in Genetic Epidemiology and Statistical Genetics. Boston, Massachusetts. February 2018.

A unified framework for composite null inference in mediation, pleiotropy, and replicability analyses

related to genetic association studies. Harvard T.H. Chan School of Public Health P01 Environmental Statistics Retreat. Boston, Massachusetts. October 2017.

The role of inflammation pathways in lung cancer and coronary artery disease. Broad Institute Statistical Genetics Meeting. Boston, Massachusetts. October 2017.

Methods for high-dimensional inference in genetic association studies. University of New Hampshire Statistics Graduate Seminar. Durham, New Hampshire. May 2017.

Set-based tests using the Generalized Berk-Jones statistic in genetic association studies. Channing Division of Network Medicine Statistical Genetics and Networks Science Meeting. Boston, Massachusetts. January 2017.

Set-based tests using the Generalized Berk-Jones statistic in genetic association studies. Broad Institute Statistical Genetics Meeting. Boston, Massachusetts. January 2017.

Pathway analysis and gene-based inference in genomic data with the Generalized Berk-Jones statistic. Harvard T.H. Chan School of Public Health P01 Environmental Statistics Retreat. Wellesley, Massachusetts. October 2016.

Testing for gene-environment interaction under misspecification of the environment. Harvard School of Public Health Program in Genetic Epidemiology and Statistical Genetics. Boston, Massachusetts. April 2016.

CONTRIBUTED TALKS AND POSTERS

A unified framework for inference in mediation, pleiotropy, and replicability analyses related to genetic association studies. Joint Statistical Meetings. Vancouver, British Columbia, Canada. August 2018.

Set-based tests using the Generalized Berk-Jones statistic in genetic association studies. Joint Statistical Meetings. Baltimore, Maryland. August 2017.

Set-based tests using the Generalized Berk-Jones statistic in genetic association studies. Eastern North Atlantic Region Spring Meeting. Washington, District of Columbia. March 2017.

Pathway analysis and gene-based inference in genomic data with the Generalized Berk-Jones statistic. Harvard T.H. Chan School of Public Health Program in Quantitative Genomics Conference. Boston, Massachusetts. November 2016.

The Generalized Berk-Jones statistic for SNP-set tests in genetic association studies. Joint Statistical Meetings. Chicago, Illinois. August 2016.

The Generalized Berk-Jones statistic for SNP-set tests in genetic association studies. Eastern North Atlantic Region Spring Meeting. Austin, Texas. March 2016.

Testing for gene-environment interaction under misspecification of the environment. Superfund Research Program Annual Meeting. San Juan, Puerto Rico. November 2015.

Testing for gene-environment interaction under misspecification of the environment. International Chinese Statistical Association - Canada Chapter Statistics and Data Science Symposium. Calgary, Alberta. August 2015.

Testing for gene-environment interaction under misspecification of the environment. Joint Statistical Meetings. Seattle, Washington. August 2015.

Professional Service	Program Committee, International Chinese Statistical Association China Conference Wuhan, 2020
	Local Committee, International Chinese Statistical Association Applied Statistics Symposium Houston, 2020
	Invited Session Organizer, Eastern North Atlantic Region Biometrics Conference Nashville, 2020
	Section Chair, Methods for Single Cell Genomic Analysis, Joint Statistical Meetings 2017
Reviewer	Annals of Applied Statistics, Bioinformatics, Biometrics, Frontiers in Genetics, Journal of the American Statistical Association, Molecular Psychiatry, npj Schizophrenia, Nucleic Acids Research Genomics and Bioinformatics, PLOS Genetics, Scientific Reports, Statistics in Medicine
Institutional Service	MD Anderson Lymphoma/Myeloma IEC Steering Committee, 2020-
	MD Anderson Scientific Review Committee, 2020-
	MD Anderson Graduate School of Biomedical Sciences Curriculum Committee, 2019-
	MD Anderson Graduate School of Biomedical Sciences Quantitative Sciences Committee, 2019-
	MD Anderson Graduate School of Biomedical Sciences Biostatistics Admissions Subcommittee, 2019-
	MD Anderson Young Faculty Committee, 2019-
	Co-organizer, Program in Quantitative Genomics Seminar, 2018-2019
	Owner and Administrator, Harvard Biostatistics Department Slack Messaging Space, 2017-2019
	Mentor, Harvard StatStart and Harvard Summer Program in Biostatistics and Computational Biology, 2015-2018

Teaching

Teaching Assistant:
- HSPH BIST245 Analysis of Multivariate and Longitudinal Data
- HSPH BIST232 Statistical Methods II (Distinction in Teaching award)
- HSPH BIST230 Probability Theory and Applications I (Distinction in Teaching award)
- HSPH BIO210 Rates and Proportions

Course Head:
- MDACC Survival Analysis, Spring 2021
- HSPH Biostatistics Summer Course in Real Analysis, Summer 2017

Selected Funding (Total Funded Effort: 90%)

NIH/NCI 5P50CA221707-02 (PI: Kopetz/Maitra)
MD Anderson Cancer Center SPORE in gastrointestinal cancer

NIH/NCI 5R01CA218230-03 (PI: Kopetz/Maitra)
Longitudinal therapeutic monitoring of colorectal cancer patients using exosome-based liquid biopsies

CPRIT RP200356 (PI: Dasari)
Activating p53 for colorectal cancer prevention

NIH/NCI 5R01CA214954-03 (PI: Wu)
Longitudinal therapeutic monitoring of colorectal cancer patients using exosome-based liquid biopsies

NIH/NCI 5R01CA196941-05 (PI: Wang)
Novel signaling pathways regulating pancreatic cancer pathogenesis

MD Anderson Breast Cancer Moon Shot (PI: Moulder)
MDACC Breast Cancer Moon Shot

NIH/NCI 5R01CA201380-05 (PI: Green)
Identifying/targeting mechanisms of lymphomagenesis driven by CREBBP inactivation

SOFTWARE
PACKAGES

GBJ: An R package for calculating the Generalized Berk-Jones statistic and its p-value. Also provides test statistic and corrected p-value for Higher Criticism, Generalized Higher Criticism, and Berk-Jones statistics when factors in a set are correlated.

GEint: An R package to calculate the exact bias of interaction coefficients in misspecified gene-environment interaction models. Also implements the Bootstrap Inference with Corrected Sandwich (BICS) procedure for testing of gene-environment interaction terms in generalized linear models.

reconstructKM: An R package to reconstruct individual-level patient data from Kaplan-Meier curves published in academic journals.

GOFexactPvalue: A C++ binary to calculate the exact p-value of Goodness-of-Fit statistics (Higher Criticism, Generalized Berk-Jones, etc.) when there are only a small number (less than 10) of correlated factors in a set.

MENTORSHIP

Jaihee Choi, Rice University Statistics PhD Research (2020-)
Carlos Vera Rocio, UTHealth GSBS PhD Thesis Committee (2020-)
Vahid Bahrambeigi, UTHealth GSBS PhD Thesis Committee (2020-)
Ziqiao Wang, UTHealth GSBS PhD Thesis Committee (2019-)
Chao Yang, UTHealth GSBS PhD Thesis Committee (2019-)
Mengting Li, Harvard Biostatistics MS Thesis Committee (2018)

PATENTS

SPLASSH: Open source software for camera-based high-speed, multispectral in-vivo optical image acquisition.