

# Tao Sun

No. 59 Zhongguancun Street  
Beijing, China 100872  
<http://taosunstat.com>

Phone: (+86)10-62515246  
Email: [sun.tao@ruc.edu.cn](mailto:sun.tao@ruc.edu.cn)  
Alt: [suntaojj@gmail.com](mailto:suntaojj@gmail.com)

## EDUCATION

---

University of Pittsburgh Ph. D. in Biostatistics Advisor: Ying Ding Committee: Wei Chen, Jong Jeong, Yu Cheng, Daniel Weeks	Pittsburgh, PA <i>April 2020</i>
Columbia University M. S. in Biostatistics	New York, NY <i>June 2015</i>
University of Pittsburgh M. S. in Immunology	Pittsburgh, PA <i>May 2013</i>
Cornell University B. S. in Animal Science <i>Distinction in Research</i>	Ithaca, NY <i>May 2009</i>

## WORK EXPERIENCE

---

Renmin University of China School of Statistics Assistant Professor of Biostatistics and Epidemiology	Beijing, China <i>August 2020 -</i>
--	--

## RESEARCH

---

My research interests focus on complex survival data with high-dimensional covariates. Specifically, I am developing statistical methods and novel applications for semiparametric methods and inference in multivariate time-to-event data (right- and interval-censored) under the high-dimensional covariate setting (e.g. genetics), machine/deep learning prediction for censored outcomes, model diagnosis for copula models, and variable selection. I also have extensive experience in analyzing high-throughput omics data (e.g. genetics, RNA sequencing, single-cell sequencing, proteomics, and metabolomics) and large-scale national health and medical records data (e.g. NHANES). In addition to methodology development, I have collaborated with researchers in the pulmonary, cancer and epidemiology for study design, data analysis and motivation of new methodology.

## AWARDS AND HONORS

---

- Best Presentation Award, Biostatistics Research Day, University of Pittsburgh, 2020.
- ASA Student of the Year, Pittsburgh Chapter, 2019.
- ENAR Distinguished Student Paper Award, 2019.
- ICSA Student Paper Award, 2019.
- LiDS Student Poster Award, 2019.
- SAMSI Travel Award for the Deep Learning Workshop, 2019.
- iBRIGHT Conference Travel Award, MD Anderson Cancer Center, 2019.
- Outstanding Teaching Assistant Award, Department of Biostatistics, University of Pittsburgh, 2019.

## FUNDING

---

- National Natural Science Foundation of China (No. 72101261). Novel Semiparametric Transformation Model under Complex Censoring, with Applications to the Progression of Chronic Diseases in the Elderly. RMB300,000 (Principal investigator, 2022-2024).
- University of Pittsburgh CTSI Quantitative Methodologies Pilot Program (NIH UL1TR001857). GWAS-based Deep Learning Survival Prediction. \$10,000 (co-Principal investigator, 2019-2020).

## PUBLICATIONS

---

1. **Tao Sun**, Yu Cheng, Ying Ding. (2022+). An Information Ratio-based Goodness-of-Fit Test for Copula Models on Censored Data. *Biometrics*. Accepted.
2. **Tao Sun**, Yunlong Li, Zhengyan Xiao, Ying Ding, Xiaojun Wang. (2022+) Semiparametric copula method for semi-competing risks data subject to interval censoring and left truncation: Application to disability in elderly. *Statistical Methods in Medical Research*. Accepted.
3. **Tao Sun**, Ying Ding. (2022+). Neural network on interval-censored data with application to the prediction of Alzheimer’s disease. *Biometrics*. Accepted.
4. Melanie J Grubisha, **Tao Sun**, Leanna Eisenman, Susan L Erickson, Shiny-yi Chou, Cassandra D Helmer, Melody Trudgen, Ying Ding, Gregg E. Homanics, Peter Penzes, Zachary P Wills, Robert A Sweet. (2021) “A Kalirin Missense Mutation Enhances Dendritic RhoA Signaling and Leads to Regression of Cortical Dendritic Arbors Across Development”. *Proceedings of the National Academy of Sciences of the United States of America*. Accepted.
5. Jun Zhang, Greg Siegle, **Tao Sun**, Wendy D’Andrea, Robert Krafty. (2021+) “Interpretable Principal Component Analysis for Multilevel Multivariate Functional Data”. *Biostatistics*. Accepted.
6. **Tao Sun**, Ying Ding. (2021) “Copula-based Semiparametric Regression Method for Bivariate Data under General Interval Censoring”. (**The earlier version won the 2019 ENAR Distinguished Student Paper Award**). *Biostatistics*. 22(2):315-330. DOI: 10.1093/biostatistics/kxz032. PMID: 31506682.
7. **Tao Sun**, Yue Wei, Wei Chen, Ying Ding. (2020) “Genome-wide association study-based Deep Learning for Survival Prediction”. (**The earlier version won the 2019 LiDS Best Poster Award**). *Statistics in Medicine*. 39(30):4605-4620. PMID: 32974946.
8. **Tao Sun**, Ying Ding. (2020) “CopulaCenR: Copula based Regression Models for Bivariate Censored Data in R”. *R Journal*. 12(1):266-282.
9. **Tao Sun**, Yi Liu, Richard J Cook, Wei Chen, Ying Ding. (2019) “Copula-based Score Test for Bivariate Time-to-event Data, with Application to a Genetic Study of AMD Progression”. *Lifetime Data Analysis*. 25(3):546-568. PMID: 30560439.
10. **Tao Sun**, Zhe Sun, Yale Jiang, Annabel Ferguson, Joseph M. Pilewski, Jay K. Kolls, Wei Chen, Kong Chen. (2019) “Transcriptomic Responses to Ivacaftor and Prediction of Ivacaftor Clinical Responsiveness”. *American Journal of Respiratory Cell and Molecular Biology*. 61(5):643-652. PMID: 30995102. (**Editorial Highlight**)
11. Ge Yang, **Tao Sun\***, Yueh-Ying Han, Franziska Rosser, Erick Forno, Wei Chen, Juan C. Celedon. (2019) “Serum cadmium and lead, wheezing and lung function in a nationwide study of adults in the United States”. *Journal of Allergy and Clinical Immunology: In Practice*. DOI: 10.1016/j.jaip.2019.05.029. PMID: 31146018. (\* co-first author)
12. Yue Wei, Yi Liu, **Tao Sun**, Wei Chen, Ying Ding. (2019) “Gene-based Association Analysis for Bivariate Time-to-event Data through Functional Regression with Copula Models”. *Biometrics*. DOI: 10.1111/biom.13165. PMID: 31625595.
13. Hehua Dai, Peixiang Lan, Daqiang Zhao, Khodor Abou-Daya, Wentao Liu, Wenhao Chen, Andrew J. Friday, Amanda L. Williams, **Tao Sun**, Jianjiao Chen, Wei Chen, Steven Mortin-Toth, Jayne S. Danska, Chris Wiebe, Peter Nickerson, Matthew L. Nicotra1, Sebastien Gingras, Hiromi Kubagawa1, Mark J. Shlomchik, Martin H. Oberbarnscheidt, Xian C. Li, Fadi G. Lakkis. (2020) “Paired immunoglobulin-like receptors mediate innate myeloid cell memory to nonself MHC molecules”. *Science*. DOI: 10.1126/science.aax4040. PMID: 32381589.
14. Hiroshi Yano, Deepali Sawant, Maria Chikina, Qianxia Zhang, Zhe Sun, **Tao Sun**, Wei Chen, Creg Workman, Dario Vignali. (2019) “Adaptive plasticity of IL10+ and IL35+ regulatory T cells and their cooperative regulation of anti-tumor immunity”. *Nature Immunology*. 20(6):724-735. PMID: 30936494.

15. Ge Yang, Yueh-Ying Han, **Tao Sun**, Ling Li, Franziska Rosser, Erick Forno, Sanjay R. Patel, Wei Chen, Juan C. Celedon. (2019) "Sleep duration, current asthma, and lung function in a nationwide study of U.S. adults". *American Journal of Respiratory and Critical Care Medicine*. 200(7):926-929. PMID: 31225970.
16. Yale Jiang, Olena Grozieva, Ting Wang, Erick Forno, Nadia Boutaoui, **Tao Sun**, Edna Acosta-Perez, Glorisa Canino, Erik Melen, Wei Chen, Juan C. Celedon. (2019). "Transcriptomics of atopy and atopic asthma in white blood cells from children and adolescents". *European Respiratory Journal*. DOI: 10.1183/13993003.00102-2019. PMID: 30923181.
17. Qi Yan, Ying Ding, Yi Liu, **Tao Sun**, Lars G Fritsche, Traci Clemons, Rinki Ratnapriya, Michael Klein, Richard Cook, Yu Liu, Ruzong Fan, Lai Wei, Goncalo Abecasis, Anand Swaroop, Emily Chew, AREDS2 Research Group, Daniel Weeks, Wei Chen. (2018). "Genome-wide Analysis of Disease Progression in Age-related Macular Degeneration". *Human Molecular Genetics*. 27(5):929-940. PMID: 29346644.
18. Kristy Boggs, Ting Wang, Abraham Orabi, Amitava Mukherjee, John Eisses, **Tao Sun**, Li Wen, Tanveer Javed, Farzad Esni, Wei Chen, Sohail Husain. (2018). "Pancreatic gene expression during recovery after pancreatitis reveals unique transcriptome profiles". *Scientific Reports*. 8(1):1406. PMID: 29362419.
19. Gabrielle Snyder, Claudia Holzman, **Tao Sun**, Marnie Bertolet, Bertha Bullen, Janet M. Catov. (2018). "Breast-feeding greater than six months is associated with smaller maternal waist circumference up to one decade after delivery". *Journal of Women's Health*. 28(4):462-472. PMID: 30481097.
20. Adam Christopher, Abraham Apfel, **Tao Sun**, Jackie Kreutzer, David Ezon. (2018). "Diastolic velocity half time is associated with aortic coarctation gradient at catheterization independent of echocardiographic and clinical blood pressure gradients". *Congenital Heart Disease*. 28(4):462-472. PMID: 30395387.
21. Sergiu Abramovici, Arun Antony, Maria Elizabeth Baldwin, Alexandra Urban, Gena Ghearing, Julie Pan, **Tao Sun**, Robert Todd Krafty, R. Mark Richardson, Anto Bagic. (2017). "Features of Simultaneous Scalp and Intracranial EEG That Predict Localization of Ictal Onset Zone". *Clinical EEG and Neuroscience*. 49(3):206-212. PMID: 29067832.
22. Abhinav P. Acharya, Kathryn M. Theisen, Andres Correa, Thiagarajan Meyyappan, Abraham Apfel, **Tao Sun**, Tatum V. Tarin, and Steven R. Little. (2017). "An inexpensive, point-of-care urine test for bladder cancer in patients undergoing hematuria evaluation". *Advanced Healthcare Materials*. 6(22):1-6. PMID: 28885787.
23. Joshua Mattila, Pauline Maiello, **Tao Sun**, Laura Via, JoAnne Flynn. (2015). "Granzyme B-expressing neutrophils correlate with bacteria load in granulomas from Mycobacterium tuberculosis-infected cynomolgus macaques". *Cellular Microbiology*. 17(8):1085-1097. PMID: 25653138.

## SOFTWARE

---

- **CopulaCenR**: a comprehensive and user-friendly R package for building copula-based regression models in bivariate censored data. Available in CRAN.
- **DNNSurv**: a Tensorflow/Keras-based R package for building and evaluating a multi-layer deep neural network survival model. A tutorial is also provided in GitHub.
- **AsthmaDiagnosis**: a shiny application for computer-aided diagnosis of children asthma; targeting for early asthma diagnosis at home or less-resourced areas.

## RESEARCH EXPERIENCE

---

- **Statistical Genetics and Computational Genomics Lab**, Pittsburgh, PA, May 2017 - present.  
Graduate Researcher – large-scale biomedical data (SNP, RNA-seq, single-cell RNA-seq, EHR, survey)  
Advisor: Wei Chen
- **Department of Biostatistics**, Pittsburgh, PA, May 2018 - August 2018.  
Graduate Researcher – proteomics and metabolomics data analyses  
Advisor: Ying Ding
- **Clinical and Translational Science Institute**, Pittsburgh, PA, Sep 2015 - April 2017.  
Statistical Consultant – study design, grant preparation, statistical analysis  
Advisor: Jong Jeong, Robert Krafty, Stephen Wisniewski
- **Columbia University**, New York, NY, Jan. 2014 - Mar 2015.  
Graduate Researcher – functional PCA analysis  
Advisor: Todd Ogden

- **University of Pittsburgh**, Pittsburgh, PA, Sep. 2009 - April. 2013.  
Immunology Researcher – adaptive immune responses to *Mycobacteria tuberculosis* infection
- **Cornell University**, Ithaca, NY, Jan. 2008 - May. 2009.  
Undergraduate Researcher – vaccine efficacy evaluation

## TEACHING EXPERIENCE

---

- **Primary Instructor.** Regression Analysis (3 credits, 80 students), 2021-2022, Renmin University of China.
- **Primary Instructor.** Survival Analysis (2 credits, 100 students), 2021-2022, Renmin University of China.
- **Primary Instructor.** Sampling Techniques (3 credits, 210 students), 2020-2022, Renmin University of China.
- **Primary Instructor.** BIOS2081: Mathematical Methods for Statistics (3 credits, 13 students), Fall 2018, University of Pittsburgh.
- **Teaching Assistant.** BIOS2054: Advanced Survival analysis, Spring 2018, University of Pittsburgh.
- **Primary Instructor.** BIOS2081: Mathematical Methods for Statistics (3 credits, 19 students), Fall 2017, University of Pittsburgh.
- **Teaching Assistant.** P8108: Survival analysis, Fall 2014, Columbia University.

## Presentations

---

1. *(Invited)* “An Information Ratio based Goodness-of-fit Test for Copula Models on Censored Data”, *The 5th International Conference on Econometrics and Statistics*, Ryukoku University, Kyoto, Japan, June 2022.
2. *(Invited)* “Neural Network on Interval Censored Data with Application to the Prediction of Alzheimer’s Disease”, *The 35th New England Statistics Symposium*, Connecticut, May 2022.
3. *(Talk)* “New Statistical Methods for Complex Survival Data with High-dimensional Covariates”, *Renmin University of China School of Statistics*, Beijing, China, September 2020.
4. *(Poster)* “Deep Learning with GWAS to Predict AMD Progression”, *iBRIGHT*, Houston, TX, November 2019.
5. *(Talk)* “GWAS-based Deep Learning for Survival Prediction”, *JSM*, Denver, CO, August 2019.
6. *(Invited)* “GWAS-based AMD Progression Using a Copula Semiparametric Model”, *ICSA Applied Statistics Symposium*, Raleigh, NC, June 2019.
7. *(Poster)* “Deep Learning with GWAS to Predict AMD Progression”, *Lifetime Data Science Annual Conference*, Pittsburgh, PA, May 2019.
8. *(Poster)* “Deep Learning with GWAS to Predict AMD Progression”, *ASA Spring Banquet*, Pittsburgh, PA, April 2019.
9. *(Talk)* “Copula-based Sieve Semiparametric Transformation Model for Bivariate Interval-censored Data”, *ENAR*, Philadelphia, PA, March 2019.
10. *(Invited)* “Asthma Risk Factor Identification and Diagnostic Prediction in a Large-scale National Survey Study”, *Children’s Hospital of Pittsburgh of UPMC*, Pittsburgh, PA, August 2018.
11. *(Poster)* “Sieve-based Copula Model for Bivariate Interval-censored Data in GWAS”, *ASA Spring Banquet*, Pittsburgh, PA, April 2018.
12. *(Invited)* “single-cell RNA data analysis in immune cell regulation”, *Department of Immunology*, Pittsburgh, PA, March 2018.
13. *(Talk)* “Copula-based Semiparametric Sieve Model for Bivariate Interval-censored Data, with an Application to Study AMD Progression”, *ENAR*, Atlanta, GA, March 2018.
14. *(Poster)* “Sieve-based Copula Model for Bivariate Interval-censored Data in GWAS”, *JSM*, Baltimore, MA, July 2017.
15. *(Poster)* “Bivariate Modeling of Genetic Effects on AMD Progression with Intermittent Assessment Times”, *ENAR*, Washington, DC, March 2017.
16. *(Poster)* “Transcriptomic Prediction of Clinical Responses and Evaluation of Responses to Ivacaftor in Cystic Fibrosis Patients”, *North American Cystic Fibrosis Conference*, Indianapolis, IN, March 2017.
17. *(Invited)* “Propensity Score Analysis of Breastfeeding Effect on Maternal Waist Circumference”, *UPMC Magee-Womens Hospital*, Pittsburgh, PA, March 2016.

## SERVICES

---

- Journal Referee (*Biometrics*, *Statistics in Medicine*, *PLOS One*, *Molecular Oncology*, *Journal of Immunology Research*)
- Lab meeting and journal club organizer, 2017-2019
- Session chairs, JSM 2019, ENAR 2019, JSM 2017
- Volunteer for LiDS conference, 2019
- Accepted student host, Department of Biostatistics, 2018-2019
- Student representative, ASA Pittsburgh Chapter, 2018-2019
- Student representative, Department of Biostatistics, 2016-2017
- Vice president, ASA Pittsburgh Student Chapter, 2016-2017

## PROGRAMMING

---

R, Python, Keras, Tensorflow, Shiny, SAS, Stata, UNIX shell scripting.

## PERSONAL

---

- Personal website: <http://taosunstat.com>
- Hobbies: reading, basketball, cooking