Yifei Sun
722 W 168th Street, R653
New York, NY 10032
ys3072@cumc.columbia.edu

Academic Appointments / Work Experience

07/2017–Present	Department of Biostatistics Mailman School of Public Health, Columbia University Assistant Professor
09/2015–06/2017	Department of Biostatistics Bloomberg School of Public Health, Johns Hopkins University Postdoctoral Fellow
09/2014–08/2015	Department of Biostatistics Bloomberg School of Public Health, Johns Hopkins University Research Assistant
09/2012–08/2014	Department of Population, Family, and Reproductive Health Bloomberg School of Public Health, Johns Hopkins University Research Assistant

Education

08/2010–08/2015	Johns Hopkins University PhD in Biostatistics, August 2015
09/2006–07/2010	Zhejiang University, Chu Kochen Honors College BS in Statistics, July 2010

Honors

07/2021	Outstanding Young Researcher Award, International Chinese Statistical Association (ICSA)
07/2020	Sanford Bolton Faculty Scholar Award, Department of Biostatistics, Columbia University Mailman School of Public Health
12/2017	Calderone Junior Faculty Award, Columbia University Mailman School of Public Health
05/2017	Student and Young Investigator Paper Award, Conference on Life Time Data Science
05/2015	Jane and Steve Dykacz Award for Outstanding Paper in Medical Statistics, Johns Hopkins University

01/2015	The Best Student Paper Award, American Statistical Association (ASA) Section on Risk Analysis Student Paper Competition
09/2011	The Glaxo SmithKline Award for best performance on PhD comprehensive exam, Johns Hopkins University

Professional Organizations, Societies, and Service

EDITORIAL SERVICE

Referee	American Journal of Epidemiology, Annals of Applied Statistics, Annals of Epidemiology, Biometrics, Biometrika, Biostatistics, BMC Medical Research Methodology, BMC Bioinformatics, Computational Statistics and Data Analysis, Electronic Journal of Statistics, Journal of Computational and Graphical Statistics, Journal of Nonparametric Statistics, Journal of the American Statistical Association (Applications & Case Studies, Theory & Methods), Journal of Applied Statistics, Journal of the Royal Statistical Society: Series C, International Journal of Biostatistics, Lifetime Data Analysis, Management Science, Pharmaceutical Statistics, Stat, Statistica Sinica, Statistical Analysis and Data Mining, Statistical Methods in Medical Research, Statistics in Biosciences, Statistics in Medicine
Associate Editor	New England Journal of Statistics in Data Science

MEMBERSHIPS AND POSITIONS

2024–2026	Treasurer, Lifetime Data Science Section, ASA
12/2022	Student Paper Award Committee, Lifetime Data Science Section, ASA
05/2021	Student Paper Award Committee, ICSA
12/2019–12/2020	Student Paper Award Committee, Lifetime Data Science Section, ASA
06/2019–07/2021	Board of Directors, ICSA
10/2014–Present	ENAR (Member)
08/2014–Present	ASA (Member)

Departmental and University Committees

08/2020–Present	Chair, Applied Qualify Exam Committee, Department of Biostatistics
08/2020–Present	Co-Chair, Curriculum Committee, Department of Biostatistics
08/2019–07/2020	Co-Chair, Applied Qualify Exam Committee, Department of Biostatistics
08/2019–Present	Member, Data Science Faculty Advisory Committee, Mailman School of Public Health
08/2018–07/2019	Organizer, Levin Lecture Series, Department of Biostatistics
07/2018–Present	Statistical Consultant, Primary Care Fellowship program, Division of General Medicine, Department of Medicine
08/2017–Present	Member, Master Admissions Committee, Department of Biostatistics
08/2017–Present	Member, Curriculum Committee, Department of Biostatistics
08/2017–Present	Member, Applied Qualify Exam Committee, Department of Biostatistics

Fellowship and Grant Support

PRESENT SUPPORT

05/2023-04/2028	1R01AG081413, NIH/NIA (Sun, Zhao) Integrative analysis for patient-centered outcomes and time-to-event data in Alzheimer's disease Role: Principle Investigator (MPI)
01/2022–12/2023	1R21HL156228, NIH/NHLBI (Sun) Dynamic and Personalized Prediction of Complex Cardiovascular Events Role: Principle Investigator
09/2019–05/2024	2U19AG033655, NIH/NIA (Albert) Biomarkers of Cognitive Decline Among Normal Individuals: The Biocard Cohort Role: Subaward PI
04/2021-03/2026	1R01HL157634 NIH/NHLBI (Oelsner) COVID-19 Lung Microvascular and Parenchymal Sequelae (Lung-MaPS) Role: Co-Investigator
02/2021-01/2024	4R00HL148511 NIH/NHLBI (Makarem) Circadian Pattern of Rest-Activity Rhythms and Blood Pressure and the Underlying Epigenetic Mechanisms Role: Co-Investigator
08/2020–07/2023	1953527, NSF (Wei) Conditional Quantile Random Forest with Biomedical and Biological Applications Role: Co-Investigator
09/2020–08/2025	1U19AG068054, NIH/NICHD & NIAID (Handen, Christian, Head, Mapstone) Alzheimer Biomarker Consortium - Down Syndrome Role: Co-Investigator
07/2019–06/2024	2R01HL093081, NIH/NHLBI (Barr) Combined Cardiopulmonary Failure in COPD: SPIROMICS HF Role: Co-Investigator
07/2018–06/2023	1K23NS105935, NIH/NINDS (Thakur) Clinical Impact of Early Pathogen Identification in Acute Neurological Infections Role: Statistician
07/1997–06/2025	5P30CA013696, NIH/NCI (Rustgi) Cancer Center Support Grant Role: Statistician

PAST SUPPORT

04/2019–03/2023	2R01HL121270, NIH/NHLBI (Barr, Laine)
	Novel Quantitative Emphysema Subtypes in MESA and SPIROMICS
	Role: Co-Investigator
08/2020-07/2022	811531 AHA (Aggarwal)
	Predictors of Weight Change During the Menopausal Transition Using a Precision Medicine Platform
10/2020-05/2022	OT2HL158276 NIH (Oelsner, Barr)
	Collaborative Cohort of Cohorts for COVID-19 Research
	Role: Co-Investigator
09/2016–08/2021	5R01AI123342 , NIH/NIAID (Zorn)
	Innate B Cell Immunity and Antibody-Mediated Rejection of Human Kidney Allografts
07/1997–06/2019	5P30CA013696, NIH/NCI (Abate-Shen)
	Cancer Center Support Grant
09/2015–06/2020	5T0BHP29302, HRSA (Shea)
	Primary Care Training and Enhancement (PCTE) Program
08/2014–04/2019	5R01HL121270, NIH/NHLBI (Barr, Laine)
	Novel Quantitative Emphysema Subtypes in MESA and Spiromics

Teaching Experience and Responsibilities

SPECIFIC COURSES

Spring 2023	Data Science II (92 enrolled students)
Spring 2022	Data Science II (102 enrolled students)
Spring 2021	Data Science II (74 enrolled students)
Spring 2020	Data Science II (58 enrolled students)
Spring 2019	Data Science II (63 enrolled students)
Spring 2018	Data Science II (54 enrolled students)

MASTER'S ADVISEES (ACADEMIC)

2022	Caroline Andy, Mei Ju Chen, Wenhao Gou, Chenxin Zhang, Ruwen Zhou
2021	Siying Li, Ruoyuan Qian, Yuanzhi Yu, Yue Lai
2020	Tian Li, Ibrahim Amidu, Jianghui Lin, Christian Pascual, Xinyao Wu
2019	Shanshan Song, Jin He, Gaeun Kim

MASTER'S ADVISEES (RESEARCH)

2023	Chaoqi Wu, Wenshan Qu, Benjamin Goebel
2022	Yatong Feng (Risk prediction for chronic obstructive pulmonary disease)
2021	Sibei Liu (Sociodemographic and clinical factors associated with etiology and clinical course in infectious and immune-mediated conditions of the nervous system)
2020	Xue Yang (Semiparametric modeling of marker trajectories)

Doctoral Examination, Advisory, and Defense Committees

2022	Madison Stoms (Oral Examination)
2022	Angel Garcia de la Garza (Oral Examination and Dissertation Defense)
2022	Yujing Yao (Oral Examination and Dissertation Defense)
2021	Yuan Chen (Oral Examination and Dissertation Defense)
2021	Patrick Hilden (Dissertation Defense)
2019	Julia Wrobel (Dissertation Defense)
2019	Huichen Zhu (Oral Examination and Dissertation Defense)

Postdoctoral Advisees

2019-2021	Ying Sheng (Jointly supervised with Chiung-Yu Huang and Mi-Ok Kim at UCSF)

Publications

* *senior authorship*
† *indicates equal contribution*
‡ *graduate student or postdoctoral fellow under my supervision*

Selected Works in Progress (submitted/in revision)

1. **Sun Y** and Sheng Y. Statistical inference for counting processes under shape heterogeneity.

2. **Sun Y**, Moghekar A, Soldan A, Pettigrew C, Greenberg B, Albert MS, Wang M-C. Cerebrospinal Alzheimer's disease biomarker patterns of change prior to the onset of mild cognitive impairment.

3. **Sun Y**, Zhao X, Chan KCG, Xu W, Allore H, Zhao Y. Semiparametric modeling of marker trajectory in the presence of left truncation and a failure event.

4. Zhu H, **Sun Y**, and Wei Y. Hybrid censored quantile regression forest to assess the heterogeneous effects.

5. Kim JS, **Sun Y**, Balte P, Cushman M, Boyle R, Tracy RP, Styer LM, Bell TM, Anderson MR, Allen NB, Schreiner P, Bowler R, Schwartz DA, Lee JS, Xanthakis V, Doyle M, Regan EA, Make BJ, Kanaya A, Coresh J, Raffield LM, Elkind M, Howard V, Ortega V, Woodruff P, Cole S, Mantis N, Parker MM, Barr RG, Oelsner EC, Demmer RT. Predictors of SARS-CoV-2 Spike 1 Antibody Response Among Vaccinated US Adults: the C4R Study.

Original, Peer Reviewed Articles

1. Sheng Y‡, **Sun Y**, McCulloch CE, Huang C-Y (2023). Scalable estimation for high velocity survival data able to accommodate adding covariates. Statistica Sinica. In press.

2. See SB, Yang X‡, Burger C, Lamarthée Snanoudj BR, Shihab R, Tsapepas DS, Roy P, Larivière-Beaudoin S, Hamelin K , Mendoza Rojas A, van Besouw NM, Bartosic A, Daniel N, Vasilescu ER, Mohan S, Cohen D, Ratner L, Baan CC, Bromberg JS , Cardinal H, Anglicheau D, **Sun Y***, Zorn E (2023). Natural antibodies are associated with rejection and long-term renal allograft loss in a multi-center international cohort. *Transplantation.* In press.

3. **Sun Y**, Chiou SH, McGarry M, Huang C-Y (2023). Dynamic risk prediction triggered by intermediate clinical events using survival tree ensembles. *Annals of Applied Statistics*. In press.

   - The work was selected to be presented in the session "The Best of AOAS" in JSM 2023

4. **Sun Y**, He X, Hu J. An omnibus test for treatment effects when many subgroups are generated via data partitioning (2022). *Annals of Applied Statistics*. 16(4): 2266-2278.

5. **Sun Y**, Chiou SH, Marr KA, Huang C-Y (2022). Statistical inference for the shape and size indexes of counting processes. *Biometrika*. 109(1):195-208.

6. Eaton A‡, **Sun Y**, Neaton J, Luo X (2022). Nonparametric estimation in an illness-death model with component-wise censoring. *Biometrics*. 78(3):1168-1180

7. Sheng Y‡, **Sun Y**, Huang C-Y, and Kim M-O (2022). Synthesizing external aggregated information in the presence of population heterogeneity: a penalized empirical likelihood approach. *Biometrics*. 78(2):679-690.

8. Clague M, Kim C, Zucker J, Green DA, **Sun Y**, Whittier S, Thakur KT (2022). Impact of implementing the cerebrospinal fluid film array meningitis/encephalitis panel on duration of intravenous acyclovir treatment. *Open Forum Infectious Diseases*. 9(8), ofac356.

9. Bhatt SP, Balte PP, Schwartz JE, Jaeger BC, Cassano PA, Chaves PH, Couper D, Jacobs Jr DR, Kalhan R, Kaplan R, Lloyd-Jones D, Newman AB, O'Connor G, Sanders JL, Smith BM, **Sun Y**, Umans JG, White WB, Yende S, Oelsner ES (2022). Pooled Cohort Probability Score for Subclinical Airflow Obstruction. *Annals of the American Thoracic Society*. 19(8):1294-1304

10. Taskiran NP, Hiura GT, Zhang X, Barr RG, Dashnaw SM, Hoffman EA, Malinsky D, Oelsner EC, Prince MR, Smith BM, Sun Y, **Sun Y**, Wild JM, Shen W, Hughes EW (2022). Mapping Alveolar Oxygen Partial Pressure in COPD Using Hyperpolarized Helium-3: The Multi-Ethnic Study of Atherosclerosis (MESA) COPD Study. *Tomography*. 8(5), 2268-2284.

11. Hermann EA, Motahari A, Hoffman EA, Allen N, Bertoni AG, Bluemke DA, Bluemke DA, Eskandari A, Gerard SE, Guo J, Hiura GT, Kaczka DW, Michos ED, Nagpal P, Pankow J, Shah S, Smith BM, Stukovsky KH, **Sun Y**, Watson K , Barr RG (2022). Pulmonary Blood Volume Among Older Adults in the Community: The MESA Lung Study. *Circulation: Cardiovascular Imaging*. 15(8), e014380.

12. Lyu T‡, Luo X, Huang C-Y, **Sun Y*** (2021). Additive rates model for recurrent event data with intermittently observed time-dependent covariates. *Statistical Methods in Medical Research*. 30(10):2239–2255.

13. Lyu T‡, Luo X, **Sun Y*** (2021). Additive-Multiplicative rates model for recurrent event data with intermittently observed time-dependent covariates. *Journal of Data Science*. 19(4):615-633.

14. Sheng Y‡, **Sun Y**, Huang C-Y, Kim M-O (2021). Synthesizing external aggregated information in the penalized Cox regression under population heterogeneity. *Statistics in Medicine*. 40(23):4915-4930.

15. Goldsmith J, **Sun Y**, Fried L, Wing J, Miller GW, and Berhane K. The emergence and future of public health data science (2021). *Public Health Reviews.* 42:1604023.

16. Jung YE, **Sun Y**, Schluger N. Impact and reach of papers posted on pre-print servers during the COVID-19 pandemic (2021). *JAMA Internal Medicine.* 181(3):395-397.

17. **Sun Y**, McCulloch CE, Marr KA, Huang C-Y (2021). Recurrent events analysis with data collected at informative clinical visits in electronic health records. *Journal of the American Statistical Association.* 116(534):594-604.

18. Ye S, Hiura G, Fleck E, Garcia A, Geleris JD, Lee P, Liyanage-Don N, Moise N, Schluger N, Singer J, Sobieszczyk M, **Sun Y**, West H, Kronish IM. (2021). Hospital Readmissions after Implementation of a Discharge Care Program for Patients with COVID-19 Illness. *Journal of General Internal Medicine.* 36(3):722-729.

19. Wysoczanski A, Angelini E, Smith BM, Hoffman E, Hiura G, **Sun Y**, Barr RG, Laine AF (2021). Unsupervised Clustering of Airway Tree Structures on High-Resolution CT: the MESA Lung Study. *ISBI' 21: IEEE International Symposium on Biomedical Imaging.*

20. **Sun Y**, Chiou SH, Wang M-C (2020). ROC-Guided survival trees and ensembles. *Biometrics.* 76(4):1177-1189.

21. Geleris J, **Sun Y**, Platt J, Zucker J, Baldwin M, Hripcsak G, Lee P, Labella A, Manson D, Kubin C, Barr RG, Sobieszczyk M, Schluger N. Observational study of hydroxychloroquine in hospitalized patients with COVID-19 (2020). *New England Journal of Medicine.* 382:2411-2418.

22. Taskiran NP, Hiura GT, Zhang X, Dashnaw SM, Hoffman EA, Malinsky D, Oelsner EC, Prince MR, Smith BM, Sun Y, **Sun Y**, Wild JM, Shen W, Barr RG, Hughes EW (2021). Estimation of the Alveolar Partial Pressure of Oxygen using Hyperpolarized Helium-3: The Multi-Ethnic Study of Atherosclerosis (MESA) COPD Study. *European Respiratory Journal.* 58: OA1566.

23. Sheng Y‡, **Sun Y**, Deng D, Huang C-Y (2020). Censored linear regression in the presence or absence of auxiliary survival information. *Biometrics.* 76(3): 734-745.

24. Namale VS, Kim C, **Sun Y**, Curcio A, Navis A, Idro R, Thakur KT (2020). Etiologies of community acquired bacterial meningitis and antibiotic resistance patterns in Africa over the last 30 years: a systematic review. Journal of Neurology & Neurophysiology, 11(6), 1-6.

25. Marr KA, **Sun Y**, Spec A, Lu N, Panackal A, Bennett J, Pappas P, Ostrander D, Datta K, Zhang SX and Williamson PR; On Behalf of the Cryptococcus Infection Network Cohort Study Working Group (2019). A multicenter, longitudinal cohort study of cryptococcosis in HIV-negative people in the United States. *Clinical Infectious Disease.* 70(2): 252-261.

26. Bai J, **Sun Y**, Schrack JA, Crainiceanu CM, Wang M-C (2018). A two-stage model for wearable device data. *Biometrics.* 74: 744-752.
    - An earlier version won the 2017 ENAR student paper Award

27. **Sun Y**, Chan G, Qin J (2018). Simple and fast overidentified rank estimation for right-censored length-biased data. *Biometrics*, 74: 77-85.

28. **Sun Y**, Qin J and Huang C-Y (2018). Missing information principle: a unified approach for general left-truncated and/or right-censored survival data problems. *Statistical Science*, 33: 261-276.

29. Shinohara RT, **Sun Y**, Wang M-C (2018). Alternating event processes during lifetimes: population dynamics and statistical inference. *Lifetime Data Analysis*. 24(1):110-125.
    - Invited submission to the special issue dedicated to Professor Jack Kalbfleisch for his contribution to survival analysis

30. **Sun Y** and Wang M-C (2017). Evaluating utility measurement with recurrent marker processes in the presence of competing terminal events. *Journal of the American Statistical Association*, 112(518): 745-756.

31. **Sun Y**, Huang C-Y, Wang M-C (2017). Nonparametric benefit-risk assessment using marker process in the presence of a terminal event. *Journal of the American Statistical Association*, 112(518): 826-836.
    - An earlier version won the Best Paper Award, 2015 ASA Section on Risk Analysis Student Paper Competition

32. Shrestha B[†], **Sun Y**[†], Faisal F, Kim V, Soares K, Blair A, Hermen JM, Narang A, Dholakia AS, Rosati L, Hacker-Prietz A, Chen L, Laheru DA, De Jesus-Acosta A, Le DT, Donehower R, Azad N, Diaz LA, Murphy A, Lee V, Fishman EK, Hruban RH, Liang T, Cameron JL, Makary M, Weiss MJ, Ahuja N, He J, Wolfgang CL, Huang C-Y, Zheng L (2017). Long-term survival benefit of upfront chemotherapy in newly diagnosed borderline resectable pancreatic cancer. *Cancer Medicine*, 6(7):1552-1562.

33. Wang M-C, **Sun Y** (2017). Nonparametric estimation of medical cost quantiles in the presence of competing terminal events. *Biostatistics & Epidemiology*, 1(1):78-91

34. Li S, **Sun Y**, Huang C-Y, Follmann DA, Krause R (2016). Recurrent event data analysis with intermittently observed time-varying covariates. *Statistics in Medicine*. 35(18): 3049-3065.
    - An earlier version won the 2016 ASA Biometrics Section Travel Award

35. Hong X, Ladd-Acosta C, Hao K, Sherwood B, Ji H, Keet CA, Kumar R, Caruso D, Liu X, Wang G, Chen Z, Ji Y, Mao G, Walker SO, Bartell TR, Ji Z, **Sun Y**, Tsai HJ, Pongracic JA, Weeks DE, Wang X (2016). Epigenome-wide association study links site-specific DNA methylation changes with cow's milk allergy *The Journal of Allergy and Clinical Immunology*, 138(3): 908-911.

36. Hong X, Hao K, Ladd-Acosta C, Hansen KD, Tsai H-J, Liu X, Xu X, Thornton TA, Caruso D, Keet CA , **Sun Y**, Wang G, Luo W, Kumar R, Fuleihan R, Singh AM, Kim JS, Story RE, Gupta RS, Gao P, Chen Z, Walker SO, Bartell TR, Beaty TH, Fallin MD, Schleimer R, Holt PG, Nadeau KC, Wood RA, Pongracic JA, Weeks DE and Wang X (2015). Genome-wide association study identifies peanut allergy-specific loci and evidence of epigenetic mediation in US children. *Nature Communications* 6. Article number: 6304.

Software

1. **Sun Y** and Chiou SH (2019). rocTree: ROC-Guided survival trees and forests. R package available on CRAN.

Invited Presentations

1. "Dynamic risk prediction triggered by intermediate events using survival tree ensembles". Department of Biotatistics, Virginia Commonwealth University, Richmond, VA. (11/2022).

2. "Dynamic risk prediction triggered by intermediate events using survival tree ensembles". Department of Statistics, University of California Irvine, Irvine, CA. (10/2022).

3. "Statistical inference for counting processes under shape heterogeneity." JSM, Washington, DC. (08/2022).

4. "Dynamic risk prediction triggered by intermediate events using survival tree ensembles". Department of Statistics, University of Connecticut, virtual. (05/2022)

5. "Dynamic risk prediction triggered by intermediate events using survival tree ensembles". Division of Biostatistics, University of Minnesota, virtual. (05/2022)

6. "Recurrent event analysis with data collected at informative clinical visits in electronic health records". Department of Statistics, North Carolina State University, virtual. (10/2021)

7. "Recurrent event trees and ensembles". ICSA 2021 Applied Statistics Symposium, virtual. (09/2021)

8. "Dynamic risk prediction using survival tree ensembles". WNAR, virtual. (06/2021)

9. "Dynamic risk prediction using survival tree ensembles". NIH/NHLBI, virtual. (03/2021)

10. "Dynamic risk prediction using survival tree ensembles". ICSA 2020 Applied Statistics Symposium, virtual. (12/2020)

11. "Statistical inference for the shape and size indexes of counting processes". Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, virtual. (10/2020)

12. "Semiparametric modeling of marker trajectory in the presence of left truncation". JSM 2020, virtual. (08/2020)

13. "ROC-guided survival trees and ensembles". Department of Epidemiology and Population Health, Albert-Einstein College of Medicine. (12/2019)

14. "Statistical inference for the shape and size indexes of counting processes". IMS-China International Conference on Statistics and Probability, Dalian, China. (07/2019).

15. "Statistical inference for the shape and size indexes of counting processes". ICSA 2019 Applied Statistics Symposium, Raleigh, NC. (06/2019)

16. "Statistical inference for the shape and size indexes of counting processes". Conference on Lifetime Data Science, Pittsburgh, PA. (05/2019)

17. "ROC-guided classification and survival trees". ICSA 2018 China Conference, Qingdao, China. (07/2018)

18. "Dealing with covariates measured at dependent follow-up visits in the estimation of Cox models". ICSA 2018 Applied Statistics Symposium, New Brunswick, NJ. (06/2018)

19. "ROC-guided classification and survival trees". IISA 2018 Conference, Gainesville, FL. (05/2018)

20. "ROC-guided classification and survival trees". Icahn School of Medicine at Mount Sinai, New York, NY. (05/2018)

21. "Dealing with covariates measured at dependent follow-up visits in the estimation of Cox models". ENAR, Atlanta, GA. (03/2018)

22. "ROC-guided classification and survival trees". Division of General Medicine, Department of Medicine, Columbia University Irving Medical Center, New York, NY. (03/2018)

23. "Rank-Based classification and survival trees". CMStatistics, London, UK. (12/2017)

24. "Missing information principle: a unified approach for general left-truncated and right-censored survival data problems". Conference on Lifetime Data Science, University of Connecticut, Storrs, CT. (05/2017)