

Shu Yang, Ph.D.

Associate Professor of Statistics
University Faculty Scholar
Statistics Department
North Carolina State University

2311 Stinson Drive
Campus Box 8203
Raleigh, NC 27695-8203
syang24@ncsu.edu

<https://shuyang.wordpress.ncsu.edu/>

Research Interests

Many important questions in chronic diseases and cancer are about the effects of treatments, e.g., approving drugs, implementing health policies, or identifying optimal personalized treatment strategies. The answers to these questions often rely on complex real-world data suffering from confounding, non-compliance, drop-outs, missing values, etc.

My research is to develop innovative statistical methods for making accurate inferences about treatment effects from complex observational and clinical studies, including marginal structural models, structural nested models, inverse probability weighting, and matching methods. This research falls into the general area of *causal inference and missing data* analyses. Our research team applies novel methods in environmental health, cardiovascular diseases, HIV infection, and cancer research to identify effective treatment strategies. I also work extensively on *high-dimensional statistics, spatial statistics, survey methodology, and data integration and fusion*.

Professional Society Memberships

Current: Caucus for Women in Statistics (CWS), American Statistical Association (ASA), Institute of Mathematical Statistics (IMS), International Chinese Statistical Association (ICSA), Western North Atlantic Region (WNAR) of the International Biometric Society, Society for Causal Inference (SCI)

Past: Eastern North Atlantic Region (ENAR) of the International Biometric Society

Employment

North Carolina State University Associate Professor in Statistics	2021–present
North Carolina State University Assistant Professor in Statistics	2016–2021
Harvard University Postdoctoral Fellow in Biostatistics Research: “Develop causal inference methods with application to initiating ART in HIV-positive patients” Adviser: Judith J. Lok, jlok@bu.edu	2014–2016
Eli Lilly and Company Intern	2014 summer

Education

Iowa State University, USA Ph.D. Co-major in Statistics and Applied Mathematics. GPA: 4.0/4.0 Thesis: “Fractional imputation methods in missing data analysis and spatial statistics” Advisers: Jae Kwang Kim, jkim@iastate.edu , and Zhengyuan Zhu, zhuz@iastate.edu	2009–2014
Beijing Normal University, P.R. China B.Sc. in Mathematics and Applied Mathematics	2005–2009

Awards & Honors

1. **Cavell Brownie Mentoring Faculty Award**, 2024, North Carolina State University, recognizes faculty members who made significant contributions to the mentoring of their colleagues and/or students.
2. **COPSS Emerging Leader Award**, 2024, Committee of Presidents of Statistical Societies, recognizes early -career statistical scientists who show evidence of and potential for leadership and who will help shape and strengthen the field.
3. **Thank an Advisor**, 2023, 2024, North Carolina State University, recognizes outstanding advisors who go the extra mile to shape students through their mentorship and dedication.
4. **Research and Innovation Seed Funding Climate Change Award**, 2023, North Carolina State University, to assist in developing innovative interdisciplinary programs.
5. **University Faculty Scholar**, 2022, North Carolina State University, recognizes and rewards emerging academic leaders and scholars among NC State’s tenured and tenure-track faculty.
6. **DRUMS Group Research Award**, 2022, Directed Research for Undergraduates in Mathematics and Statistics, North Carolina State University, awarded for teamwork and collaboration, presentation skills, and important advancement of theory, computation, and application of math and statistics.
7. **Goodnight Early Career Innovators Award**, 2021, North Carolina State University, to support early career faculty excellence and promote retention of tenure-track assistant professors whose scholarship is in STEM or STEM education.
8. **Ralph E. Powe Junior Faculty Enhancement Award**, 2018, Oak Ridge Associated Universities (ORAU).
9. **Research and Innovation Seed Funding**, 2018, North Carolina State University, to assist in developing innovative interdisciplinary programs.
10. **Thank a Teacher**, 2018, 2022, North Carolina State University, recognizes outstanding faculty going the extra mile to shape students through their teaching and mentorship.
11. **Young Investigator Scholarship**, 2016, Conference on Retroviruses and Opportunistic Infections.
12. **Harvard Postdoctoral Association Travel Award**, 2015, Harvard T. H. Chan School of Public Health, conference travel award for postdoctoral fellows.
13. **American Statistical Association (ASA) Edward C. Bryant Scholarship Award**, 2014, Westat, the award for an outstanding graduate student in survey statistics.
14. **Student Paper Competition Award**, 2014, Joint Statistical Meeting (JSM), sponsored by the Social Statistics/Government/Survey Research Methods sections of the ASA.
15. **Research Excellence Award**, 2013, Iowa State University, award for outstanding research by graduate students.
16. **Bancroft Award in Statistics**, 2012, Iowa State University, award to recognize the top student in the doctoral co-major.

Grant

Funded Research

1. [co-I] NIH-NHLBI (National Heart, Lung, and Blood Institute) grant 1R01HL169347, 2024–2029, NCSU \$261,424, role: co-I (PI: A. Zhang). Integrated detection and classification of sepsis via tensor methods using EHR.
2. [MPI] FDA (Food and Drug Administration) grant 1U01FD007934, 2023–2026, \$2,600,000 (NCSU: \$726,594), role: MPI (PIs X. Wang and S. Yang). Methods to improve efficiency and robustness of clinical trials using information from real-world data with hidden bias.
3. [PI] Eli Lilly and Company, Health Economics & Outcomes Research Pilot Funding, 2024–2024 \$40,000, role: PI. Enhancing long-term treatment effect estimation through external control borrowing.
4. [PI] NSF (National Science Foundation) grant SES 2242776, 2023–2026, \$225,000, role: PI. Causal Inference with Irregularly-spaced Observation Times.
5. [PI] NCSU Research and Innovation Seed Funding Climate Change Award, 2023–2024, \$31,250, role: PI. Harnessing Data Science to Drive Precision Policy for Marine Protected Areas.
6. [co-I] PCORI (Patient-Centered Outcomes Research Institute) grant ME-2021C3-24972, 2022–2025, NCSU \$330,560, role: co-Investigator (PI: Scharfstein, D.). Sensitivity Analysis Methods for Pragmatic Trials with Irregular and Informative Assessment Times.
7. [MPI] NIH-NIA (National Institute on Aging) grant 1R01AG066883, 2020–2024, \$1,565,763, role: MPI (PI S. Yang and X. Wang). Empower Treatment Effects Evaluation of Randomized Clinical Trials for Elderly Patients with Integrated Real-world Data.
8. [MPI] NIH-NIEHS (National Institute of Environment Health Science) grant 1R01ES031651, 2020–2024, \$1,158,927, role: MPI (PIs S. Yang and B. Reich). Spatial Causal Inference for Wildland Fire Smoke Effects on Air Pollution and Health.
9. [PI] NSF (National Science Foundation) grant DMS 1811245, 2018–2021, \$120,000, role: PI. Theory and Methods for Causal Inference in Chronic Diseases.
10. [PI] NCSU Research and Innovation Seed Funding, 2018–2019, \$31,500, role: PI. Statistical Methods for Oral Anticoagulation Therapy in Patients with Atrial Fibrillation.
11. [PI] ORAU Ralph E. Powe Junior Faculty Enhancement Award, 2018–2019, \$10,000, role: PI. Statistical Methods for Comparative Effectiveness Research in HIV infection.
12. [co-I] NCI (National Cancer Institute) grant P01 CA142538, role: co-investigator. Statistical Methods for Cancer Clinical Trials.

Funded Workshop

13. BIRS (Banff International Research Station) 5-day workshop, May 22–27, 2022, role: lead organizer, BIRS-UBC-Okanagan, “Emerging Challenges for Statistics and Data Sciences: Complex Data with Missingness, Measurement Errors, and High Dimensionality” <http://www.birs.ca/events/2022/5-day-workshops/22w5010>

Publications

* Student or intern collaborator; + corresponding author

1. L. Wu*, C. Gao*, and **S. Yang+**, B. J. Reich, and A. Rappold (2024). Estimating spatially varying health effects in app-based citizen science research, *Journal of the Royal Statistical Society: Series C*, doi:10.1093/jrssc/qlae034. [[arxiv](#)]
** Winner of the 2021 ASA Section on Statistics in Epidemiology Young Investigator Award
** Winner of the IMB Student Research Award from the 34th New England Statistics Symposium

2. J. Coulombe and **S. Yang** (2024). Multiply robust estimation of marginal structural models in observational studies subject to covariate-driven observations, *Biometrics*, **80**, ujae065. [[arxiv](#)]
3. C. Gao*, **S. Yang**, and A. Zhang (2024). Enhancing convolution neural network generalizability via low-rank weight approximation, *IET Image Processing*, 10.1049/ipr2.13205. [[arxiv](#)]
4. C. Gao*, Z. Zhang, and **S. Yang+** (2024). Causal Customer Churn Analysis with Low-rank Tensor Block Hazard Model. *ICML (41st International Conference on Machine Learning)*. [[arxiv](#)]
5. Y. Cheng* and **S. Yang** (2024). Inference for optimal linear treatment regimes in personalized decision-making. *UAI (40th Conference on Uncertainty in Artificial Intelligence)*. [[arxiv](#)]
** Selected as an oral presentation
6. T. Wang, H. Zhao*, **S. Yang+**, S. Tang, Z. Cui, L. Li, D. Faries (2024). Propensity score matching for estimating a marginal hazard ratio. *Statistics in Medicine*, doi:10.1002/sim.10103. [[arxiv](#)]
[[code](#)]
** Winner of the 2021 ENAR Distinguished Student Paper Competition Award
7. S. Fairfax* and **S. Yang** (2024). Distributional imputation for the analysis of censored recurrent events. *Statistics in Medicine*, **43**, 2622–2640. [[code](#)]
** Winner of the 2023 JSM Poster Award Competition Honorable Mention
8. B. Colnet, I. Mayer, G. Chen, A. Dieng, R. Li, G. Varoquaux, J.P. Vert, J. Josse+, **S. Yang+** (2024). Causal inference methods for combining randomized trials and observational studies: a review, *Statistical Science*, **1**, 165–191. [[arxiv](#)]
9. D. Gill, S. Lester, C. Free, A. Pfaff, E. Lversen, B. Reich, **S. Yang**, et al. (2024). A diverse portfolio of marine protected areas can better advance global conservation and equity. *Proceedings of National Academy of Sciences*, 10.1073/pnas.2313205121. (Collaborative work)
10. **S. Yang+** and X. Zhang (2024). Response to comment on “Transporting survival of an HIV clinical trial to the external target populations by Lee et al. (2024)”. *Journal of the Biopharmaceutical Statistics*, doi.org/10.1080/10543406.2024.2373449
11. D. Lee*, C. Gao*, S. Ghosh, and **S. Yang+** (2024). Transporting survival of an HIV clinical trial to the external target populations. *Journal of Biopharmaceutical Statistics*, doi.org/10.1080/10543406.2024.2330216. [[arxiv](#)] [[code](#)]
12. D. Lee*, **S. Yang**, M. Berry, T. Stinchcombe, H. Cohen, and X. Wang (2024). genRCT: A Statistical Analysis Framework for Generalizing RCT Findings to Real-World Population. *Journal of Biopharmaceutical Statistics*, doi.org/10.1080/10543406.2024.2333136. [[code](#)]
13. X. Mao, H. Wang, Z. Wang, and **S. Yang** (2024). Mixed dataframe matrix completion in survey under heterogeneous missingness. *Journal of Computation and Graphical Statistics*, doi.org/10.1080/10618600.2024.2319154. [[arxiv](#)]
14. P. Zhao*, A. Chambaz, J. Josse, and **S. Yang** (2024). Positivity-free policy learning with observational data. *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics (AISTAT)*, *PMLR* **238**:1918-1926. [[arxiv](#)]
** Top 1% selected as an oral presentation
15. S. Liu*, **S. Yang+**, Y. Zhang, and G. Liu (2024). Multiply robust estimators in longitudinal studies with missing data under control-based imputation, *Biometrics*, doi.org/10.1093/biomtc/ujad036. [[arxiv](#)]
** Winner of the 2023 ASA BIOP RISW Student Travel Award
** Winner of the 2024 ENAR RAB Student Poster Award Competition
16. Q. Guan* and **S. Yang+** (2022). A unified framework for causal inference with multiple imputation using martingale, *Statistica Sinica*, **34**, 1649–1673. [[arxiv](#)]
17. **S. Yang**, C. Gao, X. Wang, and D. Zeng (2023). Elastic integrative analysis of randomized trial and real-world data for treatment heterogeneity estimation, *Journal of the Royal Statistical Society: Series B*, **85**, 575-596. [[arxiv](#)]
18. C. Gao*, **S. Yang+**, and J. K. Kim (2023). Soft calibration for correcting selection bias under mixed-effects models, *Biometrika*, **110**, 897–911. [[arxiv](#)]

19. C. Gao* and **S. Yang** (2023). Pretest estimation in combining probability and non-probability samples, *Electronic Journal of Statistics*, **17**, 1492-1546. [[arxiv](#)]
20. J. Chu*, **S. Yang**, and W. Lu (2023). Multiply robust off-policy evaluation and learning under truncation by death, *Proceedings of the 40th (ICML) International Conference on Machine Learning*, *PMLR* **202**, 6195–6227. [[link](#)]
21. J. Chu*, W. Lu, and **S. Yang**+ (2023). Targeted optimal treatment regime learning using summary statistics, *Biometrika*, **110**, 913–931. [[arxiv](#)]
22. Y. Guan, G. L. Page, B. J. Reich, M. Ventrucchi and **S. Yang** (2023). A spectral adjustment for spatial confounding, *Biometrika*, **110**, 699–719. [[arxiv](#)]
23. M. Yu*, W. Lu, **S. Yang**, and P. Ghosh (2023). Multiplicative structural nested mean model for zero-inflated outcomes, *Biometrika*, **110**, 519–536.
24. Y. Cheng*, L. Wu, and **S. Yang** (2023). Enhancing treatment effect estimation: a model robust approach integrating randomized experiments and external controls using the double penalty integration estimator. *UAI (39th Conference on Uncertainty in Artificial Intelligence)*. [[arxiv](#)]
25. **S. Yang**, Y. Zhang, G. Liu, and Q. Guan (2023). SMIM: a unified framework of Survival sensitivity analysis using Multiple Imputation and Martingale. *Biometrics*, **79**, 230–240.
26. S. Liu*, Y. Zhang, G. T. Golm, G. Liu, and **S. Yang**+ (2023). Robust analyses for longitudinal clinical trials with dropouts and non-normal continuous outcomes. *Statistical Theory and Related Fields*, **8**, 1–14. [[arxiv](#)]
27. S. Liu*, **S. Yang**+, Y. Zhang, and G. Liu (2023). Sensitivity analysis in longitudinal clinical trials via distributional imputation, *Statistical Methods in Medical Research*, **32**, 181–194. [[arxiv](#)]
28. **S. Yang** and Y. Zhang (2023). Multiply robust matching estimators of average and quantile treatment effects, *Scandinavian Journal of Statistics*, **50**, 235–265.
29. L. Wu* and **S. Yang** (2023). Transfer learning of individualized treatment rules from experimental to real-world data, *Journal of Computation and Graphical Statistics*, **32**, 1036–1045. [[arxiv](#)]
30. E. Cho* and **S. Yang** (2023). Variable selection for doubly robust causal inference, *Statistics and Its Interface*. [[arxiv](#)]
31. D. Kong, **S. Yang**, and L. Wang (2022). Identifiability of causal effects with multiple causes and a binary outcome. *Biometrika*, **109**, 265–272.
32. Z. Jiang, **S. Yang**, and P. Ding (2022). Multiply robust estimation of causal effects under principal ignorability, *Journal of the Royal Statistical Society: Series B*, **84**, 1423–1445. [[code](#)]
33. C. Gao*, K. J. Thompson, **S. Yang** and J. K. Kim (2022). Nearest neighbor ratio imputation with incomplete multinomial outcome in survey sampling, *Journal of the Royal Statistical Society: Series A*, **185**, 1903–1930.
34. J.Y. Wang, R Wong, **S. Yang**, and G. Chan (2022). Estimation of partially conditional average treatment effect by hybrid kernel-covariate balancing, *Electronic Journal of Statistics*, **16**, 4332–4378.
35. D. Lee*, **S. Yang**+, L. Dong, X. Wang, D. Zeng, J.W. Cai (2023). Improving trial generalizability using observational studies, *Biometrics*, **79**, 1213–1225. [[code](#)]
** Winner of the 2020 ENAR Distinguished Student Paper Competition Award
36. D. Lee*, **S. Yang**+, and X. Wang (2022). Generalizable survival analysis of randomized controlled trials with observational studies, *Journal of Causal Inference*, **10**, 415–440. [[arxiv](#)] [[code](#)]
37. **S. Yang** and X. Wang (2022). RWD-integrated randomized clinical trial analysis. *2022 ASA Biopharmaceutical Report Real World Evidence* (Editors: Herbert Pang, Ling Wang, Kristi L. Griffiths), **29**, 15–21. [[Link](#)]
38. X. Mao, Z. Wang, and **S. Yang** (2022). Matrix completion under complex survey sampling, *Annals of Institute of Statistical Mathematics*, **75**, 463–492.
39. A. B. Giffin*, W. Gong, S. Majumder, A. Rappold, B. J. Reich, **S. Yang**. Estimating intervention effects on infectious disease control: the effect of community mobility reduction on Coronavirus spread, *Spatial Statistics*, doi: 10.1016/j.spasta.2022.100711. [[arxiv](#)]

40. A. B. Giffin*, B. J. Reich, **S. Yang**, and A. Rappold (2022). Generalized propensity score approach to causal inference with spatial interference, *Biometrics*, **79**, 2220–2231.
** Winner of the 2021 ENAR Distinguished Student Paper Competition
41. H. Zhao*, X. Zhang and **S. Yang** (2022). Double score matching in observational studies with multi-level treatments, *Communications in Statistics–Simulation and Computation*, doi.org/10.1080/03610918.2022.2118778.
42. H. Zhao* and **S. Yang** (2022). Outcome-adjusted balance measure for generalized propensity score model selection, *Journal of Statistical Planning and Inference*, **221**, 188–200.
** Winner of the 2021 DISS Best Poster Award
43. D. Johnson*, K. Pieper, and **S. Yang+** (2022). Treatment-specific marginal structural Cox model for the effect of treatment discontinuation, *Pharmaceutical Statistics*, **21**, 988–1004, doi:10.1002/pst.2211.
44. J. W. Yu, D. Bandyopadhyay, **S. Yang**, L. Kang, and G. Gupta (2022). Propensity score modeling in electronic health records with time-to-event endpoints: application to kidney transplantation. *Journal of Data Science*, **20**, 188–208.
45. M.Y. Huang and **S. Yang+** (2022). Robust inference of conditional average treatment effects using dimension reduction, *Statistica Sinica*, **32**, 547–567.
46. A. Larsen*, **S. Yang**, A. Rappold, and B. Reich (2022). A spatial causal analysis of wildland fire-contributed PM2.5 using numerical model output, *Annual of Applied Statistics*, **16**, 2714–2731.
47. L. Wu* and **S. Yang+** (2022). Integrative *R*-learner of heterogeneous treatment effects combining experimental and observational studies. CLEAR (*1st Conference on Causal Learning and Reasoning*), PMLR, 140, 1–S5.
48. N. Corder* and **S. Yang+** (2022). Utilizing stratified generalized propensity score matching to approximate blocked trial designs with multiple treatment levels, *Journal of Biopharmaceutical Statistics*, **32**, 373–399.
49. Y. Zhang*, **S. Yang**, W. Ye, D. E. Faries, I. Lipkovich, Z. Kadziola (2022). Best practices of double score matching for estimating causal effects, *Statistics in Medicine*, **42**, 1421–1445.
50. B. J. Reich, **S. Yang**, and Y. Guan (2022). Discussion on “Spatial+: a novel approach to spatial confounding” by Dupont, Wood and Augustin, *Biometrics*, doi:10.1111/biom.13651.
51. **S. Yang** (2022). Semiparametric efficient estimation of structural nested mean models with irregularly spaced observations, *Biometrics*, **78**, 937–949.
52. B. J. Reich, **S. Yang**, Y. Guan, A. B. Giffin, M. J. Miller and A. G. Rappold (2021). A review of spatial causal inference methods for environmental and epidemiological applications, *International Statistics Review*, **89**, 605–634.
53. **S. Yang**, J. K. Kim, and Youngdeok Hwang (2021). Integration of data from probability surveys and big found data for finite population inference using mass imputation, *Survey Methodology*, **47**, 29–58.
54. F. Cools, D. Johnson, A. J. Camm, J. P. Bassand, F. Verheugt, **S. Yang**, A. Tsiatis, D. A. Fitzmaurice, S. Z. Goldhaber, G. Kayani, S. Goto, S. Haas, F. Misselwitz, A. Turpie, K. Fox, K. Pieper, A. K. Kakkar (2021). Risks associated with discontinuation of oral anticoagulation in newly diagnosed patients with atrial fibrillation: results from the GARFIELD-AR Registry. *Journal of Thrombosis and Hemostasis*, 10.1111/jth.15415. (Collaborative work)
55. **S. Yang**, J. K. Kim, and R. Song (2020). Doubly robust inference when combining probability and non-probability samples with high-dimensional data, *Journal of the Royal Statistical Society: Series B*, **82**, 445–465.
56. **S. Yang**, K. Pieper, and F. Cools (2020). Semiparametric estimation of structural failure time model in continuous-time processes, *Biometrika*, **107**, 123–136.
57. **S. Yang** and P. Ding (2020). Combining multiple observational data sources to estimate causal effects, *Journal of American Statistical Association*, **115**, 1540–1554.

58. **S. Yang** and J. K. Kim (2020). Asymptotic theory and inference of predictive mean matching imputation using a superpopulation model framework in survey sampling. *Scandinavian Journal of Statistics*, **47**, 839–861.
59. **S. Yang** and J. K. Kim (2020). Statistical data integration in survey sampling: a review. *Japanese Journal of Statistics and Data Science*, 10.1007/s42081-020-00093-w.
60. L. Dong*, E. Laber, Y. Goldberg, R. Song, S. Yang (2020). Ascertaining properties of weighting in the estimation of optimal treatment regimes under monotone missingness, *Statistics in Medicine*, doi: 10.1002/sim.8678.
61. W. Li*, **S. Yang**+, and P. Han (2020). Robust estimation for moment condition models with data missing not at random, *Journal of Statistical Planning and Inference*, doi.org/10.1016/j.jspi.2020.01.001.
62. N. Corder* and **S. Yang** (2020). Estimating Average Treatment Effects Utilizing Fractional Imputation when Confounders are Subject to Missingness, *Journal of Causal Inference*, **8**, 249–271.
63. S. Chen, **S. Yang**, and J.K. Kim (2020). Nonparametric mass imputation for data integration. *Journal of Survey Statistics and Methodology*, doi.org/10.1093/jssam/smaa036.
64. **S. Yang**, L. Wang, and P. Ding (2019). Causal inference with confounders missing not at random, *Biometrika*, **106**, 875–888.
65. **S. Yang** (2019). Book reviews: Flexible imputation of missing data, 2nd ed. *Journal of American Statistical Association*, **114**, 1421–1421.
66. **S. Yang** and D. Zeng (2018). Discussion on penalized spline of propensity methods for treatment comparison by Zhou, Elliott and Little, *Journal of American Statistical Association*, **114**, 30–32.
67. **S. Yang** and J. J. Lok (2018). Sensitivity analysis for unmeasured confounding in coarse structural nested mean models, *Statistica Sinica*, **28**, 1703–723.
68. **S. Yang** (2018). Propensity score weighting for causal inference with clustered data, *Journal of Causal Inference*, doi.org/10.1515/jci-2017-0027.
69. **S. Yang** and J. K. Kim (2018). Nearest neighbor imputation for general parameter estimation in survey sampling, *Advances in Econometrics*, **39**, 211–236.
70. **S. Yang** and P. Ding (2018). Asymptotic inference of causal effects with observational studies trimmed by the estimated propensity scores, *Biometrika*, **105**, 487–493.
71. Z. Wang, J. K. Kim, and **S. Yang** (2018). An approximate Bayesian inference under informative sampling, *Biometrika*, **105**, 91–102.
72. J. J. Lok, **S. Yang**, B. Sharkey, Hughes, M (2018). Estimation of the cumulative incidence function under multiple dependent and independent censoring mechanisms, *Lifetime Data Analysis*, **24**, 201–223.
73. **S. Yang**, A. A. Tsiatis, and M. Blazing (2018). Modeling survival distribution as a function of time to treatment discontinuation: a dynamic treatment regime approach, *Biometrics*, **74**, 900–909.
74. **S. Yang** and J. K. Kim (2017). A semiparametric inference to regression analysis with missing covariates in survey data, *Statistica Sinica*, **27**, 261–285.
75. J. K. Kim and **S. Yang** (2017). A note on multiple imputation under complex sampling, *Biometrika*, **104**, 221–228.
76. **S. Yang** and J. K. Kim (2017). Discussion: dissecting multiple imputation from a multi-phase inference perspective: what happens when god's, imputer's and analyst's models are uncongenial? by X. Xie and X. L. Meng, *Statistica Sinica*, **27**, 1568–1573.
77. **S. Yang**, and J. J. Lok (2016). A goodness-of-fit test for structural nested mean models, *Biometrika*, **103**, 734–741.
78. **S. Yang**, and J. K. Kim (2016). Fractional imputation in survey sampling: a comparative review, *Statistical Science*, **31**, 415–432.
79. **S. Yang**, G. Imbens, Z. Cui, D. Faries and Z. Kadziola (2016), Propensity score matching and stratification in observational studies with multi-level treatments, *Biometrics*, **72**, 1055–1065.

80. **S. Yang** and J. K. Kim (2016). A note on multiple imputation for method of moments estimation, *Biometrika*, **103**, 244–251.
81. **S. Yang** and J. K. Kim (2015). Likelihood-based inference with missing data under missing-at-random, *Scandinavian Journal of Statistics*, **43**, 436–454.
** Winner of the 2014 JSM Student Paper Competition Award
82. K. L. Peyer, G. Welk, L. B. Davis, **S. Yang**, and J. K. Kim (2015). Factors associated with parent concern for child weight and parenting behaviors, *Childhood Obesity*, **11**, 269–274. (Collaborative work)
83. **S. Yang** and Z. Zhu (2015). Variance estimation and kriging prediction for a class of non-stationary spatial models, *Statistica Sinica*, **25**, 135–149.
84. J. K. Kim and **S. Yang** (2014). Fractional hot deck imputation for robust estimation under item nonresponse in survey sampling, *Survey Methodology*, **40**, 211–230.
85. J. K. Kim, Z. Zhu, and **S. Yang** (2013). Improved estimation for June Area Survey incorporating several information, *Proceedings 59th ISI World Statistics Congress, Hong Kong, China*, 199–04.
86. **S. Yang**, J. K. Kim and D. W. Shin (2013). Imputation methods for quantile estimation under missing at random, *Statistics and Its Interface*, **6**, 369–377.
87. **S. Yang**, J. K. Kim and Z. Zhu (2013). Parametric fractional imputation for mixed models with nonignorable missing data, *Statistics and Its Interface*, **6**, 339–347.

Technical Reports

88. **S. Yang** and P. Ding. Two-phase rejective sampling. *Journal of the Royal Statistical Society: Series B*, revision. [[arxiv](#)]
89. P. Sang, D. Kong, and **S. Yang**+. Functional principal component analysis for longitudinal observations with sampling at random, *Biometrika*, revision. [[arxiv](#)]
90. C. Gao*, A.R. Zhang, and **S. Yang**+. Causal inference on sequential treatments via tensor completion. *Journal of the Royal Statistical Society: Series B*, revision submitted.
91. C. Gao*, **S. Yang**+, M. Shan, W. Ye, I. Lipkovich, and D. Faries. Integrating randomized trial data with external controls: a semiparametric approach with selective borrowing, *Biometrika*, revision submitted. [[arxiv](#)] [[code](#)]
** Winner of the 2024 ICSA Student Paper Award
92. C. Gao*, **S. Yang**, M. Shan, W. Ye, and I. Lipkovich. Doubly protected estimation for survival outcomes utilizing external controls for randomized clinical trials, *Journal of American Statistical Association*.
93. C. Gao*, X. Zhang, and **S. Yang**. Omnibus sensitivity analysis of externally controlled trials with intercurrent events, *Biometrics*, submitted. [[arxiv](#)]
94. G. Mao, **S. Yang**+, and X. Wang. Statistical inference for heterogeneous treatment effect with right-censored data from synthesizing randomized clinical trials and real-world data. *Biometrika*, submitted.
95. B. Smith*, Y. Gao*, **S. Yang**, A. Apter, and D. Scharfstein. Trials with irregular and informative assessment times: a sensitivity analysis approach, *Biometrics*, revision submitted. [[arxiv](#)]
96. P. Zhao*, J. Josse, and **S. Yang**. Efficient and robust transfer learning of optimal individualized treatment regimes with right-censored survival data. *Journal of Machine Learning Research*, major revision. [[arxiv](#)]
97. **S. Yang** and Z. Zhu. Semiparametric estimation of spectral density and variogram with irregular observations. *Journal of Statistical Planning and Inference*, revision. [[arxiv](#)]
98. **S. Yang**, S. Liu, D. Zeng, X. Wang. Improved inference for heterogeneous treatment effects using real-world data subject to hidden confounding, *Bournoulli*, revision submitted. [[arxiv](#)]
99. A. B. Giffin*, B. J. Reich, **S. Yang**, and A. Rappold. Instrumental variables, spatial confounding and interference, *Journal of Causal Inference*, submitted. [[arxiv](#)]
100. Y. Zhang*, D. Kong, and **S. Yang**+. Towards R-learner of conditional average treatment effects

with a continuous treatment: T-identification, estimation, and inference. *Biometrika*, submitted. [[arxiv](#)]

**** Winner of the 2023 ASA Section on Nonparametric Statistics Student Paper Award**

101. Y. Zhang* and **S. Yang+**. Semiparametric localized principal stratification analysis with continuous strata. *Journal of the Royal Statistical Society: Series B*, submitted. [[arxiv](#)]
102. S. Xu*, **S. Yang**, B. J. Reich. A Bayesian non-parametric method for estimating causal quantile effects. *Journal of Computation and Graphical Statistics*, submitted.
103. X. Tan, **S. Yang**, W. Ye, D. E. Faries, I. Lipkovich, Z. Kadziola. When doubly robust methods meet machine learning for estimating treatment effects from real-world data: a comparative study. *Pharmaceutical Statistics*, submitted. [[arxiv](#)]
104. T. Hong*, W. Lu, **S. Yang**, and P. Ghosh. Multivariate choice models with irregularly spaced longitudinal observations: application to the lockdown effect on consumer behaviors, *Biometrics*, submitted. [[arxiv](#)]
105. T. Hong*, W. Lu, **S. Yang**, and P. Ghosh. Multivariate zero-inflated causal model for regional mobility restriction effects on consumer spending, *Journal of American Statistical Association*, submitted.
106. D. Faries, C. Gao, X. Zhang, C. Hazlett, J. Stamey, **S. Yang**, et al. Real effect or bias? Best practices for evaluating the robustness of real-world evidence through quantitative sensitivity analysis for unmeasured confounding, *Pharmaceutical Statistics*, revision. [[Authorea](#)] [[arxiv](#)]
107. Y. Fang, P. Mishra-Kalyani, X. Zhang, S. Gruber, **S. Yang**, P. Ding, M. Shan, J.Y. Lee, M. van der Laan, D. Faries, and H. Lee. Sensitivity analysis for unmeasured confounding in medical product development and evaluation using real-world evidence, *Statistics in Biopharmaceutical Research*, submitted. [[arxiv](#)]
108. Z. Wang, **S. Yang** and J.K. Kim. Multiple bias-calibration for adjusting selection bias of non-probability samples using data integration. *Journal of Official Statistics*, revision. [[arxiv](#)]
109. Z. Chen* and **S. Yang+**. Enhancing generalization and transportation for survival data with the minimal covariates. *Journal of the Royal Statistical Society: Series A*, revision.
110. K. Jiang, X. Lai, **S. Yang**, Y. Gao and X. Zhou. A practical analysis procedure on generalizing comparative effectiveness in randomized clinical trial to the real-world trial-eligible population. *Journal of the Biopharmaceutical Statistics*, submitted. [[arxiv](#)]
111. K. Jiang, W. Hu, **S. Yang**, X. Lai and X. Zhou. Improve sensitivity analysis synthesizing randomized clinical trials with limited overlap. *Biometrika*, submitted. [[arxiv](#)]
112. X. Chen, W. Lu, **S. Yang**, and D. Bandyopadhyay. Off-policy evaluation with irregularly-spaced and outcome-dependent observation process. *Journal of American Statistical Association*, submitted. [[arxiv](#)]
113. G. Wang*, S. Liu*, and **S. Yang**. Continuous-time structural failure time model for intermittent treatment. [[arxiv](#)]
114. L. Neas, W. Steinhardt, KL Hill, R Short, E. Hubal, BJ Reich, **S. Yang**, A. Sheng, and AG Rappold. Fluorosurfactants in groundwater increase the incidence of chronic health conditions among California Medicare beneficiaries. *Environmental Health Perspectives*, submitted. [[MedRxiv](#)]
115. X. Ye, **S. Yang+**, X. Wang, and Y. Liu. Integrative analysis of high-dimensional RCT and RWD subject to censoring and hidden confounding. *Journal of American Statistical Association*, submitted.
116. J. Chu*, **S. Yang**, W. Lu, and P. Ghosh. A shadow variable approach to policy evaluation and learning with one-sided feedback. *NeurIPS*, submitted.
117. Y. Xu, **S. Yang**, and J. Kang. Bayesian structured mediation analysis with unobserved confounders. *NeurIPS*, submitted. [[arxiv](#)]
118. H. Zhao*, T. Wang, **S. Yang**, Z. Cui, I. Lipkovich, and D. Fairs. Propensity score matching for estimation of pairwise marginal hazard ratios. *Communications in Statistics – Theory and Methods*, revision.

119. Y. Liu*, Y. Wang, and **S. Yang**. Rate doubly robust estimation for weighted average treatment effects. *Statistics in Medicine*, major revision.
120. C. Cui*, Y. Zhang*, **S. Yang**, B. J. Reich, and D. Gill. Matching estimators of causal effects in clustered observational studies with application to quantifying the impact of marine protected areas on biodiversity, *Biometrics*, submitted. [[arxiv](#)]
121. T. Xu*, Y. Zhang*, and **S. Yang**. Augmented match weighted estimators for average treatment effects. *JRSSA*, submitted. [[code](#)]
122. Y. Cheng* and S. Yang. Bagged bias-corrected matching estimator for average treatment effects. *Scandinavian Journal of Statistics*, submitted.

Thesis

S. Yang (2014). Fractional imputation method of handling missing data and spatial statistics. Iowa State University. [[Link](#)]

Presentations

Atlantic/Pacific Causal Inference Conference (A/PCIC); Eastern North American Region of International Biometric Society Spring Meeting (ENAR); International Chinese Statistical Association (ICSA); Lifetime Data Science Conference (LiDS); Joint Conference on Statistics and Data Science (JCSDS); Joint Statistical Meeting (JSM); International Conference on Statistics and Data Science (ICSDS)

Short courses and workshops

1. Statistical methods for time to event data from multiple sources: a causal inference perspective, *Half-Day Short Course, ENAR, New Orleans, LA, USA. March 23–26, 2024*
2. Statistical methods for time-to-event data from multiple sources: a causal inference framework, *Full-Day Short Course, ICSA, Nashville, TN, USA. June 16, 2023*
3. Design and analysis of randomized clinical trials with real-world data using causal inference framework and Bayesian methods, *Half-Day Short course, ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop, Rockville, MD, USA. September 27–29, 2023*
4. Unveiling the Power of Real-World Data: A Causal Inference Framework for Designing and Analyzing Randomized Clinical Trials, *2nd CANSSI-NISS sponsored Health Data Science Workshop, Half-Day Short Course, Waterloo, Ontario, Canada. August 2, 2023*
5. Statistical methods for time-to-event data from multiple sources: a causal inference framework, *Full-Day Short Course, LiDS, Raleigh, NC, USA. May 31, 2023*
6. Design and analysis of randomized clinical trials with real-world data, *Half-Day Preconference Workshop on the Society of Clinical Trial, Baltimore, MD, USA. May 21–24, 2023*
7. Improved causal inference combining randomized clinical trials and observational studies, *Half-day lectures, Canadian Statistical Sciences Institute (Collaborative Research Teams) Summer School, University of Ottawa, Ontario, Canada. (Invited). July 5–8, 2022*

Seminars and talks

1. Methods to improve efficiency and robustness of clinical trials using information from real-world data with hidden bias. *ASA BIOP DL Webinar: Real-World Evidence HHSU01 FDA grantees series Part 2, Virtual. (Invited by Herb Pang) August 2024*
2. Mitigating bias in treatment effect estimation: strategies for utilizing external controls in randomized trials. *JSM, Portland, Oregon. (Invited by Jiwei Zhao) August 2024*
3. Causal Customer Churn Analysis with Low-rank Tensor Block Hazard Model. *ICML, Vienna, Austria. (Poster presentation) July 2024*

4. Multiply robust off-policy evaluation and learning under truncation by death. *JCSDS (Joint Conference on Statistics and Data Science in China)*, Kunming, Yunnan, China. (Invited by Jae Kwang Kim) July 2024
5. Multiply robust off-policy evaluation and learning under truncation by death. *PCIC*, Shanghai, Zhejiang, China. (Invited) July 2024
6. Multiply robust off-policy evaluation and learning under truncation by death. *ICSA*, Nashville, TN, USA (Invited by Yingqi Zhao) June 2024
7. Multiply robust off-policy evaluation and learning under truncation by death. *University of Paris Cite*, Paris, France (Invited) May 2024
8. Multiply robust off-policy evaluation and learning under truncation by death. *Inria*, Montpellier, USA, May 2024
9. Dagstuhl Seminar 24202 on Causal Inference for Spatial Data Analytics, *Schloss Dagstuhl, Leibniz-Zentrum für Informatik*, Wadern, Germany, (Invited) May 2024
10. Strategies for utilizing external controls in randomized trials while mitigating bias in treatment effect estimation. *The 7th Stat4Onc Annual Symposium*, UConn, Storrs, Connecticut, USA. (Invited), May 2024
11. Test-Then-Pool: A uniformly valid inferential framework for data integration, *Colloquium Speaker*, University of Virginia, VA, USA. (Invited), March 2024
12. Enhancing treatment effect estimation: a model robust approach integrating randomized experiments and external controls using the double penalty integration estimator. *ENAR* Baltimore, MD, USA. (Invited) March 2024
13. Test-Then-Pool: A uniformly valid inferential framework for data integration, *Colloquium Speaker*, University of Texas at Dallas, TN, USA. (Invited) March 2024
14. Test-Then-Pool: A uniformly valid inferential framework for data integration, *2024 Statistics Annual Winter Workshop*, University of Florida, Gainesville, FL, USA. (Invited) January 2024
15. [Mitigating bias in treatment effect estimation: strategies for utilizing external controls in randomized trials](#), *Workshop "Extending Inference to a New Target Population" at ICERM (Institute for Computational and Experimental Research in Mathematics)*, Brown University, Providence, USA. (Invited) November 2023
16. [Discussion of "Entropy balancing vs vector-based kernel weighting for causal inference in categorical treatment settings" by Feyman, Lum, Asfaw and Garrido](#), *Annual Health Econometrics Workshop*, University of Southern California in Los Angeles, LA, USA. (Invited) October 2023
17. Robust sensitivity analysis in clinical trial studies with missing data under control-based imputation, *ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop*, Rockville, MD, USA. (Invited) September 2023
18. Design and analysis of randomized clinical trials with real-world data using causal inference framework and Bayesian methods, *Half-Day Short course*, *ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop*, Rockville, MD, USA. September 2023
19. Enhancing Treatment Effect Estimation: A Model Robust Approach Integrating Randomized Experiments and Historical Controls using the Double Penalty Integration Estimator. *PCIC*, Beijing and virtual meeting, China. (Invited) September 2023
20. Should double score matching replace propensity score matching in comparative observational research? *JSM*, Toronto, ON, Canada (Invited) August 2023
21. Efficient and robust transfer learning of optimal individualized treatment regime with right-censored survival data. *JSM*, Toronto, ON, Canada. (Invited) August 2023
22. Test-Then-Pool: a uniformly valid inferential framework for data integration. *2nd CANSSI-NISS sponsored Health Data Science Workshop*, University of Waterloo, Waterloo, ON, Canada. (Invited) August 2023

23. Unveiling the Power of Real-World Data: A Causal Inference Framework for Designing and Analyzing Randomized Clinical Trials. *2nd CANSSI-NISS sponsored Health Data Science Workshop, University of Waterloo, Waterloo, ON, Canada. (Invited Short Course) August 2023*
24. Test-Then-Pool: a uniformly valid inferential framework for data integration. *JCSDS, Peking University, Beijing, China. (Invited) July 2023*
25. Doubly robust estimators for generalizing treatment effects on survival outcomes from randomized controlled trials to a target population. *Wuhan University, Wuhan, China. (Invited) July 2023*
26. Test-Then-Pool: a uniformly valid inferential framework for data integration. *Zhongnan University of Economics and Law, Wuhan, China. (Invited) July 2023*
27. Targeted optimal treatment regime learning. *Central China Normal University, Wuhan, China. (Invited) July 2023*
28. Doubly robust estimators for generalizing treatment effects on survival outcomes from randomized controlled trials to a target population. *ICSA-China Chapter, Chengdu, Sichuan, China. (Invited) June 2023*
29. Doubly robust estimators for generalizing treatment effects on survival outcomes from randomized controlled trials to a target population, *LiDS, Raleigh, NC (Invited) June 2023*
30. Statistical methods for time-to-event data from multiple sources: a causal inference framework, *Full-Day Workshop, LiDS, Raleigh, NC, USA. May 2023*
31. Target optimal Treatment Regime Learning, *ACIC, Austin, TX, USA. May 2023*
32. Design and analysis of randomized clinical trials with real-world data, *Half-Day Preconference Workshop on the Society of Clinical Trial, Baltimore, MD, USA. May 2023*
33. Causal inference methods for combining randomized trials and real-world data, *Nankai University, Online Seminar Series, China. (Invited) May 2023*
34. Empower RCT analysis with integrated information from RWD, [Clubear](#) Online Seminar Series, *China. (Invited) April 2023*
35. Empower RCT analysis with integrated information from RWD, *Colloquium Speaker, University of Toronto, Toronto, ON, Canada. (Invited) April 2023*
36. Empower RCT analysis with integrated information from RWD, *Colloquium Speaker, University of Waterloo, Waterloo, ON, Canada. (Invited) April 2023*
37. Generalizable survival analysis of randomized clinical trials with observational studies, *ENAR Nashville, Tennessee, USA. (Invited) March 2023*
38. Empower RCT analysis with integrated information from RWD, *Colloquium Speaker, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. (Invited) March 2023*
39. Empower RCT analysis with integrated information from RWD, *Biostatistics Colloquium Speaker, University of Columbia University, New York City, New York, USA. (Invited) March 2023*
40. Empower RCT analysis with integrated information from RWD, *Causal Inference Seminar, Boston University, Boston, MA, USA. (Invited) February 2023*
41. Empower RCT analysis with integrated information from RWD, *Seminar, University of Maryland School of Medicine, Virtual, USA. (Invited) February 2023*
42. Targeted optimal treatment regime learning using summary statistics, *Causal Inference Research Group Seminar, University of North Carolina, Chapel Hill, NC, USA. (Invited) November 2022*
43. Estimating spatially varying health effects in app-based citizen science research, *North Carolina SOT (Society of Toxicology) Fall Meeting on Climate Change Exposures, North Carolina Central University, Raleigh, NC, USA. (Invited) October 2022*
44. Targeted optimal treatment regime learning using summary statistics. *Virtual Colloquium Speaker, Renmin University, Beijing, China. (Invited) October 2022*
45. Targeted optimal treatment regime learning using summary statistics. *Colloquium Speaker, University of Michigan, Ann Arbor, MI, USA. (Invited) October 2022*
46. Future of Statistics. *Iowa State University, 75th Anniversary Celebration Conference, Ames, IA, USA. (Invited Panelist) October 2022*

47. Targeted optimal treatment regime learning using summary statistics. *CSSM seminar, Iowa State University, Ames, IA, USA. (Invited) September 2022*
48. A unified framework for DSM: theory, balance measure, and practice. *ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop, Rockville, MD, USA. (Invited) September 2022*
49. Spatial Causal inference methods for environmental and epidemiological applications. *JSM, Washington, DC, USA. (Invited) August 2022*
50. Optimal treatment regime estimation for a target population with summary statistics. *JSM, Washington, DC, USA. (Invited) August 2022*
51. Generalizable survival analysis of randomized clinical trials with observational studies. *ICSA-Canada Chapter, Banff, Alberta, Canada. (Invited). July 2022*
52. Elastic integrative analysis of the heterogeneity of treatment effect combining randomized trial and observational data. *Society for Epidemiologic Research Conference, Chicago, USA. (Invited). June 2022*
53. Generalizable survival analysis of randomized clinical trials with observational studies. Workshop “[Missing Data and Survival Analysis](#)”, *Angers, France. (Invited) May 2022*
54. Test-based integrative estimator for heterogeneous treatment effect combining randomized trials and real-world data. *Online Causal Inference Seminar, Stanford, CA, USA. (Invited) April 2022*
55. Estimating spatially varying health effects in app-based citizen science research. *ENAR, Houston, TX, USA. (Invited) March 2022*
56. Causal inference, missing data analysis, and data integration. *Iowa State University (STATERS), Ames, IA, USA. (Invited) March 2022*
57. Estimating spatially varying health effects in app-based citizen science research. *Colloquium Speaker, Chinese University of Hong Kong, Hong Kong, China. (Invited) March 2022*
58. Estimating spatially varying health effects in app-based citizen science research. *Colloquium Speaker, University of Missouri, Columbia, USA. (Invited) February 2022*
59. Discussion on “Nonrandom samples and causal inference” by Daniel McCaffrey. *Discussant, Online Causal Inference Seminar, Stanford, USA. (Invited) January 2022*
60. Continuous-time causal models with irregularly spaced longitudinal observations. *Online Colloquium Speaker, Sichuan University, China. (Invited) October 2021*
61. Causal inference with irregularly spaced longitudinal observations. *Online Colloquium Speaker, University of Utah, USA. (Invited) October 2021*
62. Continuous-time causal models with irregularly spaced longitudinal observations. *Online Colloquium Speaker, University of Pittsburg, USA. (Invited) October 2021*
63. Empower trial analysis with real world data. *Chancellor visit. NC State, USA. October 2021*
64. Data integration: a new paradigm for survey statistics. *NESS-New England Statistical Society. University of Rhode Island, USA. (Invited) October 2021*
65. Continuous-time causal models with irregularly spaced longitudinal observations. *Online Colloquium Speaker, University of Waterloo, Canada. (Invited) September 2021*
66. Empower RCT analysis with integrated information from RWD. *Online Colloquium Speaker, Wuhan University, Wuhan, China. (Invited) September 2021*
67. Statistical methods for improving randomized clinical trial analysis with integrated information from real-world evidence studies. *ICSA, virtual meeting, USA. (Invited) September 2021*
68. Semiparametric efficient estimation of structural nested mean models with irregularly spaced observations. *PCIC, virtual meeting, China. (Invited) September 2021*
69. Data integration: a new paradigm for survey statistics. Keynote talk. *Baltic-Nordic-Ukrainian Network on Survey Statistics – 2021 Summer School. (Invited) September 2021*
70. Improved inference for heterogeneous treatment effects using real-world data subject to hidden confounding. *JSM, virtual meeting, USA. (Invited) August 2021*
71. Statistical methods for improving RCT analysis with integrated information from RWE. *DISS (Duke Industry Statistics Symposium), virtual meeting, USA. (Invited) April 2021*
72. Improved inference for heterogeneous treatment effects using real-world data subject to hidden

- confounding. *UC Berkeley Causal Inference Group, virtual meeting, USA. (Invited) March 2021*
73. Improved inference for heterogeneous treatment effects using real-world data subject to hidden confounding. *Online Colloquium Speaker, University of Minnesota, USA. (Invited) February 2021*
 74. Multiply robust matching estimators of average and quantile treatment effects. *Real-World Analytics Forum, Eli Lilly and Company, virtual meeting, USA. (Invited) February 2021*
 75. Improved inference for heterogeneous treatment effects using real-world data subject to hidden confounding. *CFE-CMStatistics Conference, virtual meeting, King's College London, UK. (Invited) December 2020*
 76. Data Integration and Causal Inference. *Online Colloquium Speaker, University of McGill, Canada. (Invited) October 2020*
 77. Test-based integrative analysis of randomized clinical trial and real-world data. *Online Colloquium Speaker, Peking University, China. (Invited) October 2020*
 78. Test-based integrative analysis of randomized clinical trial and real-world data. *Online Colloquium Speaker, University of Louisville, USA. (Invited) October 2020*
 79. Improved inference for heterogeneous treatment effects using real-world data subject to hidden confounding. *Pacific Causal Inference Conference, virtual meeting, Beijing, China. (Invited) September 2020*
 80. Integrative analysis of randomized clinical trial and real-world data. *JSM, virtual meeting, USA. (Invited) August 2020*
 81. Elastic integrative analysis of randomized trial and real-world data for heterogeneity of treatment effect. *Online Seminar Speaker, Fudan University, Shanghai, China. (Invited) June 2020*
 82. Doubly robust inference when combining probability and nonprobability samples with high-dimensional data. *ENAR, virtual meeting, USA. (Invited) March 2020*
 83. Double score matching estimators of average treatment effects and quantile treatment effects. *Colloquium Speaker, Iowa State University, Ames, USA. (Invited) November 2019*
 84. Integrative analysis of randomized clinical trial and real-world evidence studies. *Colloquium Speaker, Texas A&M University, Texas, USA. (Invited) October 2019*
 85. Semiparametric Estimation of Continuous-Time Structural Failure Time Model. *JSM, Denver, Colorado, USA. (Invited) July 2019*
 86. JASA, Applications and Case Studies, discussant, *JSM, Denver, Colorado, USA. (Invited) July 2019*
 87. Integrative analysis of randomized clinical trial with real world evidence studies. *ICSA, Tianjin, Hebei, China. (Invited) July 2019*
 88. Integrative analysis of randomized clinical trial with real world evidence studies. *Real-World Analytics, Eli Lilly and Company, Online Meeting, USA. (Invited) June 2019*
 89. Causal inference with confounders missing not at random. *ICSA, Raleigh, NC, USA. (Invited) June 2019*
 90. Causal inference with confounders missing not at random. *Departmental Seminar Speaker, NC State, Raleigh, NC, USA. April 2019*
 91. Causal inference with confounders missing not at random. *ENAR, Philadelphia, PA, USA. (Invited) March 2019*
 92. Causal inference with confounders missing not at random. *Colloquium Speaker, Baylor University, Waco TX, USA. (Invited) February 2019*
 93. Propensity score matching and subclassification in observational studies with multi-level treatments. *The International Biometrics Society Journal Club. Webinar (Invited) December 2018*
 94. Asymptotic inference of causal effects with observational studies trimmed by the estimated propensity score. *ICSA, New Brunswick, NJ, USA. (Invited) June 2018*
 95. Combining multiple observational data sources to estimate causal effects. *ACIC, Pittsburgh, PA, USA. (Invited) May 2018*
 96. Dynamic regime marginal structural models to survival distribution as a function of time to treatment discontinuation. *ENAR, Atlanta, GA, USA. (Invited) March 2018*

97. Dynamic regime marginal structural models to survival distribution as a function of time to treatment discontinuation. *Colloquium Seminar Speaker, Kansas State University, Manhattan, KS, USA. March 2018*
98. Modeling survival distribution as a function of time to treatment discontinuation. *Departmental Seminar Speaker, NC State University, Raleigh, NC, USA. September 2017*
99. Modeling survival distribution as a function of time to treatment discontinuation. *Colloquium Speaker, Purdue University, West Lafayette, IN, USA. (Invited) September 2017*
100. Estimation of the cumulative incidence function under multiple dependent and independent censoring mechanisms. *JSM, Baltimore, MD, USA. (Invited) August 2017*
101. Nonparametric identification of causal effects with confounders subject to instrumental missingness. *ACIC, Chapel Hill, NC, USA. (Poster) May 2017*
102. Propensity score weighting for causal inference with multi-stage data. *ENAR, Washington, DC, USA. (Invited Poster) March 2017*
103. A note on multiple imputation of handling missing data under complex sampling. *JSM, Chicago, IL, USA. (Invited) August 2016*
104. Estimation and goodness-of-fit test of structural nested mean models. *ENAR, Austin, TX, USA. March 2016*
105. Optimal estimation of coarse structural nested mean models with application to initiating HAART in HIV-positive patients. *National Institute of Health (NIH) Infectious Disease Research: Quantitative Methods and Models in the era of Big Data Statistical Workshop, Bethesda, MD, USA. (poster) Nov 2015*
106. Double robust goodness-of-fit test of coarse structural nested mean models with application to initiating HAART in HIV-positive patients. *ACIC, Philadelphia, PA, USA. May 2015*
107. Optimal estimation of coarse structural nested mean models. *ENAR, Miami, FL, USA. March 2015*
108. Fractional imputation method for missing data analysis: a review. *Workshop on Analyzing Complex Survey Data with Missing Item Values, National Institute of Statistical Sciences (NISS), Washington D.C., USA. (Invited) October 2014*
109. Likelihood-based inference with missing data under missing-at-random. *JSM, Boston, MA, USA. August 2014*
110. Likelihood-based inference with missing data under missing-at-random. *ICSA-KISS symposium, Portland, OR, USA. June 2014*
111. Propensity score matching and subclassification with multivalued treatments. *ACIC, Providence, RI, USA. May 2014*
112. Fractional hot deck imputation for robust estimation under item nonresponse in survey sampling. *JSM, Montreal, Canada. August 2013*
113. Parametric fractional imputation for longitudinal data with non-ignorable missing data. *JSM, San Diego, CA, USA. August 2012*
114. Quantify uncertainty in image classification. *National Agricultural Statistics Service (NASS), Washington, DC, USA. May 2011*

Students and Postdoc Supervision

Current Postdoc fellows

1. Ke Zhu (PhD, Tsinghua University, co-supervised with Xiaofei Wang)

Current Ph.D. students

2. Yi Liu (PhD Advisor)
3. Yujing Gao (PhD Advisor)

4. Sarah Reigel (PhD Advisor)
5. Siqi Cao (PhD Advisor)
6. Shubhajit Sen (PhD Co-Advisor with Rui Song)
7. Xiaodan Zhou (PhD Co-Advisor with Brian Reich)
8. Sihyung Park (PhD Co-Advisor with Wenbin Lu)

Past Postdoc fellows (with the first place of employment)

9. Hyemin Yeon (PhD, Iowa State University, co-supervised with Brain Reich, 2023–2024; Kent State University)

Past Ph.D. students (with the first place of employment)

Dissertation with chapters in causal inference

10. Yuwen Cheng (PhD Advisor, 2019–2024; Amgen)
Advanced statistical causal inference and individualized treatment regime learning: data integration, inference, and matching [[link](#)]
11. Jianing Chu (PhD Co-Advisor with Wenbin Lu, 2019–2024; Amazon)
Advances in policy evaluation and learning: targeting, truncation by death, and one-sided feedback [[link](#)]
12. Chenyin Gao (PhD Advisor, 2019–2024; Harvard)
Advanced statistical methods for data integration and tensor completion in causal inference [[link](#)]
13. Siyi Liu (PhD Advisor, 2019–2024; Merck)
Treatment effect evaluation in longitudinal studies with missing data and data integration [[link](#)]
14. Taekwon Hong (PhD Co-Advisor with Wenbin Lu, 2019–2024; FDA)
Analysis of Irregularly Spaced Longitudinal Market Transaction Data [[link](#)]
15. Tanchumin Xu (PhD committee member, 2018–2023, Wells Fargo)
Advances in causal inference and the study of interlocus gene conversion [[link](#)]
16. Yichi Zhang (PhD Co-Advisor with Minh Tang, 2018–2023; Indiana University Bloomington)
Statistical inference with randomized SVD for signal-plus-noise matrix models and causal inference with continuous interventions [[link](#)]
17. Yunshu Zhang (PhD Advisor, 2018–2023; U Penn)
Doubly robust estimators of causal effects in observational studies: theory and practice [[link](#)]
18. Joe Zhao (PhD Co-Advisor with Emily Hector, 2018–2023; SAS Institute)
Advances in Matching Methods for Causal Inference with Multiple Treatments [[link](#)]
19. Eunah Cho (PhD Advisor, 2016–2022; LG Display)
Robust Causal Inference Methods for Using Randomized Clinical Trial and Observational Study [[link](#)]
20. Dasom Lee (PhD Co-Advisor with Sujit Ghosh, 2017–2022; United Therapeutics)
Semiparametric Inference of Randomized Controlled Trials and Observational Studies [[link](#)]
21. Steven Xu (PhD Co-Advisor with Brian Reich, 2017–2022; DoorDash)
Advances in Semiparametric Quantile Regression [[link](#)]
22. Lili Wu (PhD Advisor, 2016–2021, Microsoft)
Spatially Varying and Multi-Source Data Integrative Causal Inference [[link](#)]
23. Nathan Corder (PhD Co-Advisor with Marie Davidian, 2014–2021, Eli Lilly)
Conducting Causal Inference on Partially Observed Data via Imputation and Matching [[link](#)]
24. Andrew Giffin (PhD Co-Advisor with Brian Reich, 2016–2020; FDA)
Methods for Causal Inference on Spatial Data with Environmental and Public Health Applications [[link](#)]
25. Lin Dong (PhD Co-Advisor with Eric Laber, 2015–2019; Google)
Semiparametric Methods for Decision Making and Causal Effect Generalization [[link](#)]
26. Yiming Wang (PhD Committee Member, 2017–2022; Pinterest)

- Covariance Function Estimation and Causal Inference Methods* [[link](#)]
27. Can Cui (PhD Committee Member, 2017–2021; Eli Lilly)
Advanced Methods in Bayesian Variable Selection and Causal Inference [[link](#)]
28. Miao Yu (PhD Committee Member, 2017–2021; Meta)
Online Testing and Semiparametric Estimation of Complex Treatment Effect [[link](#)]
29. Alexandra Larsen (PhD Committee Member, 2015–2019; EPA)
Spatial Methods for Quantifying the Impact of Wildfire Smoke on Air Quality in the U.S. [[link](#)]
30. Qian Guan (PhD Committee Member, 2015–2019; Meta)
Bayesian Methods for Optimal Treatment Allocation and Causal Inference [[link](#)]

PhD committee member for:

NCSU

Jake Koerner (advised my Ana-Maria Staicu, 2021–2025), Yang (Alina) Xu (advised by Rui Song and Wenbin Lu, 2021–2025), Dongjae Son (advised by Brian Reich and Erin Schliep), Richard Watson (Causal inference and causal graphs), Ye Shen, Yuhe Dale Gao, Han Wang, Charlie Song, Xuan Liu, Tanchumin Xu, Peter Norwood, Nicholas Larsen, Jesse Clifton, Jianian Wang, Jingwen Zhao (Math department), Mingqi Wu (PhD Examiner, McGill University), Hengrui Cai (PhD, 2017–2022), Breanna Swan (PhD, 2016–2021), Eric Rose (PhD, 2015–2019)

University of Copenhagen

Zehao Su (advised by Frank Eriksson, Biostatistics, University of Copenhagen)

Research mentor for undergraduate students

31. Anthony Wang (NCSU, Research Mentor, 2022 Summer REU program, currently PhD candidate in UNC)
32. Kelly Wentzlof (Indiana University, Research Mentor, 2022 Summer REU program)
33. Johnny Rajala (University of Maryland, Research Mentor, 2022 Summer REU program)
34. Miontranese Green (California State University, Research Mentor, 2022 Summer REU program)

Student Awards

Advisor of the student winners

1. **ICSA Student Paper Award**, ICSA Applied Statistics Symposium, Chenyin Gao, 2024
2. **ENAR RAB Student Poster Award Competition**, 2nd place, Eastern North American Region of International Biometric Society Spring Meeting, Siyi Liu, 2024
3. **FDA-OCE-ASA Oncology Educational Fellowship**, Taekwon Hong, 2023
4. **JSM Student Poster Award (honorable mention)** from the ASA Biopharmaceutical Section in cooperation with the FDA Statistical Association, Sarah Fairfax, 2023
5. **BIOP RISW Student Travel Award** from the ASA Biopharmaceutical Section in cooperation with the FDA Statistical Association, Siyi Liu, 2023
6. **JSM Student Paper Award** from the ASA Section on Nonparametric Statistics, Joint Statistical Meetings, Yichi Zhang, 2023
7. **IBM Student Research Award**, the 34th New England Statistics Symposium, Lili Wu, 2021
8. **DISS Best Poster Award** from Duke Industry Statistics Symposium, Joe Zhao, 2021
9. **JSM Paper Award** from the Section on Statistics in Epidemiology, Lili Wu, 2021
10. **ENAR Distinguished Paper Competition Award**, Eastern North American Region of International Biometric Society Spring Meeting, Andrew Giffin, 2021
11. **ENAR Distinguished Paper Competition Award**, Eastern North American Region of International Biometric Society Spring Meeting, Shuhan Tang, 2021
12. **ENAR Distinguished Paper Competition Award**, Eastern North American Region of International Biometric Society Spring Meeting, Lin Dong, 2020

Editorial Services

1. Associate Editor, *Biometrics* 2018–2026
2. Associate Editor, *Journal of Computational and Graphical Statistics* 2021–present
3. Associate Editor, *Journal of the Royal Statistical Society Series A* 2023–2027
4. Guest co-editor, *Journal of Agricultural Biological and Environmental Statistics*
Special Issue on Causal Inference 2021–2023
5. Associate Guest Editor, *Statistica Sinica* 2016–2018

6. NSF Advisory Panel for the MMS program 2021–2022
7. NSF DMS panel review 2019/2022
8. NSF MMS panel review 2021
9. Natural Sciences and Engineering Research Council of Canada grant review 2021
10. MRC (Medical Research Council, UKRI) grant review 2021
11. NIH/NHLBI Clinical Trial Review Study Section (CLTR) 2023
12. HEI (Health Effects Institute) review panel 2023–present
13. NIH/NIDA CTN Data Safety Monitoring Board 2024–present
14. Reviewer for *Annals of Applied Statistics*, *Biometrika*, *Biometrical Journal*, *Canadian Journal of Statistics*, *Communications in Statistics*, *Epidemiologic Methods*, *Electronic Journal of Statistics*, *International Journal of Biostatistics*, *Journal of Business & Economic Statistics*, *Journal of Multivariate Analysis*, *Journal of the American Statistical Association*, *Journal of the Korean Statistical Society*, *Journal of the Royal Statistical Society: Series B*, *Statistica Sinica*, *Stat, Statistics and Its Interface*, *Survey Methodology*, *NeurIPS*, *ICML*

Professional Experience

1. Program committee, 1st ACM SIGSPATIAL International Workshop on Spatiotemporal Causal Analysis (STCausal Workshop) 2024
2. ASA BIOP RISW Steering Committee 2024–present
3. Scientific committee, the PCIC in Shanghai, China 2024
4. Awards Committee of the ASA Statistics In Epidemiology Career Awards 2024
5. Junior research award committee, the ICSA China Conference in Wuhan 2024
6. Scientific committee, the 2024 ICSA symposium at Vanderbilt 2024
7. Consultant for NIH grant R21 (PI: Lu Tang), Federated learning methods for heterogeneous and distributed Medicaid data 2023–2025
8. Academic member of ASA BIOP RWE Scientific Working Group “Sensitivity Analysis” team. 2022–2024
9. Academic leader of ASA BIOP RWE Scientific Working Group “Sensitivity Analysis” team. 2024–present
10. Full Member of the CHHE (Center for Human Health and the Environment). 2022–present
11. Executive committee, the DISS (Duke Industry Statistics Symposium) in Durham 2021
12. SAMSI faculty fellow for the Causal Inference program 2020
13. Executive committee, the ICSA symposium in Raleigh 2019
14. Treasurer, the ICSA symposium in Raleigh 2019
15. Translated chapters in Book “Causal Inference for Statistics, Social and Biomedical Sciences: An Introduction” into Chinese 2015–2017
16. Organized invited sessions at conferences

- “Advances in statistical methods for biological data and health informatics” at *Ecostat, Beijing Normal University, Beijing, China, July 2024*
- “Statistical and causal inference with irregularly spaced observation times” at *ICSA, Nashville, TN, June 2024*
- “Advancing clinical trial analysis: integrating external data and novel estimation techniques” at *ICSA, Nashville, TN, June 2024*
- “Mitigating Incomplete Data Biases – A Modern Take” at *ICSA-China Chapter, Chengdu, Sichuan, China, July 2023*
- “Recent Advances in Causal Inference and Missing Data Analysis” at *ICSA-Canada Chapter, Banff, Alberta, Canada, July 2022*
- "Sensitivity Analysis with Nonignorable Missing Data: Recent Work from Academia, Industry, and Regulatory Agencies" at *JSM, virtual meeting. August 2021*
- "Integrated analysis of RCTs and Population-based Studies" at *DISS, virtual meeting. April 2021*
- "Matching Methods for Causal Inference with Emerging Data and Statistical Challenges" at *JSM, virtual meeting. August 2020*
- "Integrative analysis of clinical trials and real-world evidence studies" at *ENAR, virtual meeting. March 2020*
- “Recent Advance of Causal Inference in Failure Time Settings” at *JSM, Colorado, Denver, USA. August 2019*
- “Causal Inference and Missing Data Analysis: Identification and Estimation” at *ICSA, Raleigh, NC, USA. June 2019*
- “Causal Inference with Non-ignorable Missing Data: New Developments in Identification and Estimation” at *ENAR, Philadelphia, PA, USA. March 2019*
- “Statistical Inference in Air Pollution and Health Epidemiology” at *ICSA, New Brunswick, NJ, USA. June 2018*
- “Causal Inference and Data Fusion” at *ACIC, Pittsburgh, PA, USA. May 2018*
- “Causal Inference for Continuous-time Processes: New Developments” at *ENAR, Washington, DC, USA. March 2017*
- “New Developments in Structural Nested Models with Medical Applications” at *ACIC, Philadelphia, PA, USA. May 2015*
- “Big Data Techniques for Survey Data Integration” at *JSM, Seattle, WA, USA. August 2015*

Software

R packages for matching

1. [multilevelMatching](#) for implementing a novel matching procedure to compare multiple treatments simultaneously from the observational data. ([CRAN](#))
2. [dsmatch](#) implements double score matching for average treatment effect and quantile treatment effect estimation.

R packages for continuous-time causal inference

3. [contTimeCausal](#) provides estimation methods for continuous-time structural failure time models (ctSFTM) and continuous-time Cox marginal structural models (ctCoxMSM) ([CRAN](#)).

R packages for improving the external validity of randomized trials

4. [genRCT](#) generalizes the average treatment effect (ATE) from the trial using observational studies (OS) for binary/continuous/survival outcomes.

R packages for integrative analysis

5. [IntegrativeFPM](#) implements integrative analyses for the finite population mean combining probability and non-probability samples with high-dimensional data.
6. [IntegrativeCI](#) implements integrative analyses for the average treatment effect combining big main data and smaller validation data.
7. [IntegrativeHTE](#) implements elastic analyses for the heterogenous treatment effects combining trials and real-world data.
8. [IntegrativeHTEcf](#) implements integrative analyses for the heterogenous treatment effects combining a randomized trial and confounded real-world data.
9. [intRlearner](#) implements an integrative R-learning analysis of the heterogenous treatment effects by combining a randomized trial and confounded real-world data using the notation of confounding function.

R packages for missing data and causal inference

10. [miATE](#) implements a unified bootstrap inference of the average treatment effect after multiple imputation. based on martingales.
11. [pace](#) implements various estimators of principal strata average causal effects from observational studies.
12. [smim](#) implements Survival sensitivity analysis using *Multiple Imputation and Martingale*.
13. [RealWorld-DoublyRobustML](#) implements various doubly robust estimators using machine learned nuisance functions, including DSM, AIPW, TMLE, and PENCOMP.

Teaching (enrollment numbers)

1. ST371 Introduction to Probability and Distribution Theory
2016 Fall Section 001 (100), Section 002 (87)
2018 Spring Section 001 (90), Section 002 (85)
2. ST520 Statistical Principles of Clinical Trials
2018 Fall (26)
2019 Spring (27)
3. ST790 Causal Inference
2023 Fall (27)

Departmental Service

1. Student admission committee (2024)
2. Certified facilitator of workshops “Entering Mentoring” (2023)
3. Faculty member of the Building Future Faculty program (2023)
4. Faculty search committee (2022, Chair)
5. Biostatistics working group (2022, 2023, 2024 Chair)
6. Diversity committee (2020–2022)
7. Junior faculty representative (2019, 2020)
8. Faculty search committee (2019)
9. Beach trip committee (2019)
10. Department seminar committee (2018, 2021 Chair)
11. Qualify exam committee (2017)