Shu Yang, Ph.D.

North Carolina State University
E-mail: syang24@ncsu.edu
Website: https://shuyang.wordpress.ncsu.edu/

Statistics Department
2311 Stinson Drive
Campus Box 8203
Raleigh, NC 27695-8203

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 the novel methods in environmental health, cardiovascular diseases, HIV infection and cancer research to identify effective treatment strategies.

Employment

North Carolina State University 2016–presence
Assistant Professor in Statistics

Merck & Co 2021–presence
Consultant

Eli Lilly and Company 2019–presence
Consultant

Harvard University 2014–2016
Postdoctoral Fellow in Biostatistics
Research: "Develop causal inference methods with application to initiating ART in HIV-positive patients"
Adviser: Judith J. Lok, jlok@hsph.harvard.edu

Eli Lilly and Company 2014
Intern

Education

Iowa State University, USA 2009 Aug –2014 May
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

Beijing Normal University, P.R. China 2005 Sep–2009 July
B.Sc. in Mathematics and Applied Mathematics

Awards & Honors

1. **Goodnight Early Career Innovators Award,** 2021, North Carolina State University, to support early career faculty excellence and promotes retention of tenure-track assistant professors whose scholarship is in STEM or STEM education.
2. **Best Poster Award** from Duke Industry Statistics Symposium (DISS), 2021.
3. **Joint Statistical Meetings Paper Award** from the Section on Statistics in Epidemiology, 2021.
4. **Distinguished Paper Competition Award** from Eastern North American Region of International Biometric Society Spring Meeting, 2021.
5. **Distinguished Paper Competition Award** from Eastern North American Region of International Biometric Society Spring Meeting, 2020.
6. **Ralph E. Powe Junior Faculty Enhancement Award**, 2018, Oak Ridge Associated Universities (ORAU).
7. **Research and Innovation Seed Funding**, 2018, North Carolina State University, to assist in developing innovative interdisciplinary programs.
8. **Young Investigator Scholarship**, 2016, Conference on Retroviruses and Opportunistic Infections.
9. **Harvard Postdoctoral Association Travel Award**, 2015, Harvard T. H. Chan School of Public Health, conference travel award for postdoctoral fellows.
10. **American Statistical Association (ASA) Edward C. Bryant Scholarship Award**, 2014, Westat, award for an outstanding graduate student in survey statistics.
11. **Student Paper Competition Award**, 2014, Joint Statistical Meeting (JSM), sponsored by the Social Statistics/Government/Survey Research Methods sections of the ASA.
12. **Research Excellence Award**, 2013, Iowa State University, award for outstanding research by graduate students.
13. **Bancroft Award in Statistics**, 2012, Iowa State University, award to recognize the top student in the doctoral co-major.

Grant

Funded Workshop

1. BIRS (Banff International Research Station) 5-day workshop, 2022 BIRS-UBC-Okanagan, "Emerging Challenges for Statistics and Data Sciences: Complex Data with Missingness, Measurement Errors, and High Dimensionality"
2. 2020-2021 SAMSI (Statistical and Applied Mathematical Sciences Institute) workshop for Program on Data Science in the Social and Behavioral Sciences

Funded Research

3. NIH-NIA (National Institute on Aging) grant 1R01AG066883, 2020–2024, $1,565,763, role: PI. Empower Treatment Effects Evaluation of Randomized Clinical Trials for Elderly Patients with Integrated Real-World Data.
4. NIH-NIEHS (National Institute of Environment Health Science) grant 1R01ES031651, 2020–2024, $1,158,927, role: PI. Spatial Causal Inference for Wildland Fire Smoke Effects on Air Pollution and Health.
5. NSF (National Science Foundation) grant DMS 1811245, 2018–2021, $120,000, role: PI. Theory and Methods for Causal Inference in Chronic Diseases.
6. NCSU Research and Innovation Seed Funding, 2018–2019, $31,500, role: PI. Statistical Methods for Oral Anticoagulation Therapy in Patients with Atrial Fibrillation.
7. ORAU Ralph E. Powe Junior Faculty Enhancement Award, 2018–2019, $10,000, role: PI. Statistical Methods for Comparative Effectiveness Research in HIV infection.

8. NCI (National Cancer Institute) grant P01 CA142538, role: co-Investigator. Statistical Methods for Cancer Clinical Trials.

Submitted

9. NSF (National Science Foundation, PI: Dong, J.Y.) 2022–2025, role: co-PI. SCH: Integrating Wearable Multimodal Sensor System with Artificial Intelligence for Precision Diagnosis of Respiratory Diseases.
10. PCORI (Patient-Centered Outcomes Research Institute, PI: Scharfstein, D.) 2021–2023, role: co-Investigator. Sensitivity Analysis Methods for Pragmatic Trials with Irregular and Informative Assessment Times.

Publications

* Student or intern collaborator; + corresponding author

1. 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*, doi: 10.1111/insr.12452.
2. **S. Yang** (2021). Semiparametric efficient estimation of structural nested mean models with irregularly spaced observations, *Biometrics*, doi.org/10.1111/biom.13471.
3. M.Y. Huang and **S. Yang**+ (2021). Robust inference of conditional average treatment effects using dimension reduction, *Statistica Sinica*, doi:10.5705/ss.202020.0409.
4. D. Kong, **S. Yang**, and L. Wang (2021). Muti-cause causal inference with unmeasured confounding and binary outcome, *Biometrika*, in press.
5. **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.
6. 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)
7. **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.
8. **S. Yang**, K. Pieper, and F. Cools (2020). Semiparametric estimation of structural failure time model in continuous-time processes, *Biometrika*, **107**, 123-136.
9. **S. Yang** and P. Ding (2020). Combining multiple observational data sources to estimate causal effects, *Journal of American Statistical Association*, **115**, 1540-1554.
10. **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.
11. **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.
12. 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.
13. 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.
14. 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.

15. 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.
16. **S. Yang**, L. Wang, and P. Ding (2019). Causal inference with confounders missing not at random, *Biometrika*, **106**, 875-888.
17. **S. Yang** (2019). Book reviews: Flexible imputation of missing data, 2nd ed. *Journal of American Statistical Association*, **114**, 1421-1421.
18. **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.
19. **S. Yang** and J. J. Lok (2018). Sensitivity analysis for unmeasured confounding in coarse structural nested mean models, *Statistica Sinica*, **28**, 1703-1723.
20. **S. Yang** (2018). Propensity score weighting for causal inference with clustered data, *Journal of Causal Inference*, doi.org/10.1515/jci-2017-0027.
21. **S. Yang** and J. K. Kim (2018). Nearest neighbor imputation for general parameter estimation in survey sampling, *Advances in Econometrics*, **39**, 211-236.
22. **S. Yang** and P. Ding (2018). Asymptotic inference of causal effects with observational studies trimmed by the estimated propensity scores, *Biometrika*, **105**, 487-493.
23. Z. Wang, J. K. Kim, and **S. Yang** (2018). An approximate Bayesian inference under informative sampling, *Biometrika,* **105**, 91-102.
24. 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.
25. **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.
26. **S. Yang** and J. K. Kim (2017). A semiparametric inference to regression analysis with missing covariates in survey data, *Statistica Sinica*, **27**, 261-285.
27. J. K. Kim and **S. Yang** (2017). A note on multiple imputation under complex sampling, *Biometrika,* **104**, 221-228.
28. **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.
29. **S. Yang**, and J. J. Lok (2016). A goodness-of-fit test for structural nested mean models, *Biometrika,* **103**, 734-741.
30. **S. Yang**, and J. K. Kim (2016). Fractional imputation in survey sampling: a comparative review, *Statistical Science*, **31**, 415-432.
31. **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. With R package available "multilevelMatching".
32. **S. Yang** and J. K. Kim (2016). A note on multiple imputation for method of moments estimation, *Biometrika*, **103**, 244-251.
33. **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
34. 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)
35. **S. Yang** and Z. Zhu (2015). Variance estimation and kriging prediction for a class of non-stationary spatial models, *Statistica Sinica*, **25**, 135-149.

36. 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.
37. 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*, 199-204.
38. **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.
39. **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

40. Q. Guan* and **S. Yang+**. A unified framework for causal inference with multiple imputation using martingale. [arxiv]
41. L. Dong*, **S. Yang+,** X. Wang, D. Zeng, J.W. Cai. Integrative analysis of randomized clinical trials with real world evidence studies. [arxiv]** Winner of the 2020 ENAR Distinguished Student Paper Competition Award
42. **S. Yang**, Y. Zhang, G. Liu, and Q. Guan. SMIM: a unified framework of *S*urvival sensitivity analysis using *M*ultiple *I*mputation and *M*artingale. [arxiv]
43. S. Tang*, **S. Yang+**, T. Wang, Z. Cui. L. Li, D. Faries. Causal inference of hazard ratio based on propensity score matching. [arxiv]** Winner of the 2021 ENAR Distinguished Student Paper Competition Award
44. A. B. Giffin*, B. J. Reich , **S. Yang**+, and A. Rappold. Generalized propensity score approach to causal inference with spatial interference. [arxiv]** Winner of the 2021 ENAR Distinguished Student Paper Competition
45. X. Mao, Z. Wang, and **S. Yang**. Matrix completion for survey data prediction with multivariate missingness. [arxiv]
46. **S. Yang** and Z. Zhu. Semiparametric estimation of spectral density and variogram with irregular observations. [arxiv]
47. **S. Yang**, D. Zeng, X. Wang. Improved inference for heterogeneous treatment effects using real-world data subject to hidden confounding. [arxiv]
48. **S. Yang**, X. Wang, and D. Zeng. Elastic integrative analysis of randomized trial and real-world data for treatment heterogeneity estimation. [arxiv]
49. B. Colnet, I. Mayer, G. Chen, A. Dieng, R. Li, G. Varoquaux, J.P. Vert, J. Josse+, **S. Yang**+. Causal inference methods for combining randomized trials and observational studies: a review. [arxiv]
50. **S. Yang** and Y. Zhang. Multiply robust matching estimators of average and quantile treatment effects. [arxiv]
51. Y. Zhang*, **S. Yang**, W. Ye, Douglas E. Faries, I. Lipkovich, Z. Kadziola. Best practices of double score matching for estimating causal effects.
52. Y. Guan, G. L. Page, B. J. Reich, M. Ventrucci and **S. Yang**. A spectral adjustment for spatial confounding. [arxiv]
53. L. Wu*, and **S. Yang+**, B. J. Reich, and A. Rappold. Estimating spatially varying health effects in app-based citizen science research. [arxiv]** Winner of the 2021 ASA Section on Statistics in Epidemiology Young Investigator Award
54. L. Wu* and **S. Yang**. Transfer learning of individualized treatment rules from experimental to real-world data.
55. A. Larsen*, **S. Yang**, A. Rappold, and B. Reich. A spatial causal analysis of wildland fire-contributed PM2.5 using numerical model output. [arxiv]
56. D. Johnson, K. Pieper, and **S. Yang**+. Treatment-specific marginal structural Cox model for the effect of treatment discontinuation.

57. J.Y. Wang, R Wong, **S. Yang**, and G. Chan. Estimation of partially conditional average treatment effect by hybrid kernel-covariate balancing. [arxiv]
58. A. B. Giffin*, B. J. Reich, **S. Yang**, and A. Rappold. Instrumental variables, spatial confounding and interference. [arxiv]
59. 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. [arxiv]
60. Z. Jiang, **S. Yang**, and P. Ding. Multiply robust estimation of causal effects under principal ignorability.
61. M. Yu*, W. Lu, **S. Yang**, and P. Ghosh. Multiplicative structural nested mean model for zero-inflated outcomes.
62. C. Gao* and **S. Yang**. Pretest estimation in combining probability and non-probability samples.
63. C. Gao*, K. J. Thompson, **S. Yang** and J. K. Kim. Nearest neighbor ratio imputation with incomplete multinomial outcome in survey sampling.
64. E. Cho* and **S. Yang**. Variable selection for doubly robust causal inference.
65. H. Zhao* and **S. Yang.** Outcome-adjusted balance measure for generalized propensity score model selection.** Winner of the 2021 DISS Best Poster Award
66. S. Liu*, **S.Yang**+, Y. Zhang, and G. Liu. Sensitivity analysis in longitudinal clinical trials via distributional imputation.
67. B. J. Reich, **S. Yang,** and Y. Guan. Discussion on "Spatial+: a novel approach to spatial confounding" by Dupont, Wood and Augustin.

Thesis

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

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 integrative analysis

4. IntegrativeFPM implements integrative analyses for the finite population mean combining probability and non-probability samples with high-dimensional data.
5. IntegrativeCI implements integrative analyses for the average treatment effect combining big main data and smaller validation data.
6. IntegrativeHTE implements elastic analyses for the heterogenous treatment effects combining trials and real-world data.
7. IntegrativeHTEcf implements integrative analyses for the heterogenous treatment effects combining a randomized trial and confounded real-world data.

R packages for missing data and causal inference

8. miATE implements a unified bootstrap inference of the average treatment effect after multiple imputation. based on martingales.
9. pace implements various estimators of principal strata average causal effects from observational studies.
10. smim implements *S*urvival sensitivity analysis using *M*ultiple *I*mputation and *M*artingale.

Presentations

Atlantic/Pacific Causal Inference Conference (A/PCIC); Eastern North American Region of International Biometric Society Spring Meeting (ENAR); International Chinese Statistical Association (ICSA); Joint Statistical Meeting (JSM)

1. TBD. Workshop "Missing Data and Survival Analysis", *Angers, France. (Invited) May 30--June 1, 2022*
2. Estimating spatially varying health effects in app-based citizen science research. *ENAR (Invited) 2022*
3. TBD. *Online Colloquium Speaker, University of Utah, USA. (Invited) 2021*
4. TBD. *Online Colloquium Speaker, University of Pittsburg, USA. (Invited) October 2021*
5. Statistical methods for improving randomized clinical trial analysis with integrated information from real-world evidence studies. *ICSA (Invited) September 12--15*
6. Data integration: a new paradigm for survey statistics. Keynote talk. *Baltic-Nordic-Ukrainian Network on Survey Statistics – 2021 Summer School. (Invited) September 3 2021*
7. Semiparametric efficient estimation of structural nested mean models with irregularly spaced observations. *PCIC, virtual meeting, China. (Invited) September 1 2021*
8. Improved inference for heterogeneous treatment effects using real-world data subject to hidden confounding. *JSM, virtual meeting, USA. (Invited) August 2021*
9. Statistical methods for improving RCT analysis with integrated information from RWE. *DISS (Duke Industry Statistics Symposium), virtual meeting, USA. (Invited) April 2021*
10. 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*
11. Improved inference for heterogeneous treatment effects using real-world data subject to hidden confounding. *Online Colloquium Speaker, University of Minnesota, USA. (Invited) February 2021*
12. Multiply robust matching estimators of average and quantile treatment effects. *Real-World Analytics Forum, Eli Lilly and Company, virtual meeting, USA. (Invited) February 2021*
13. 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*
14. Data Integration and Causal Inference. *Online Colloquium Speaker, University of McGill, Canada. (Invited) October 2020*
15. Test-based integrative analysis of randomized clinical trial and real-world data. *Online Colloquium Speaker, Peking University, China. (Invited) October 2020*
16. Test-based integrative analysis of randomized clinical trial and real-world data. *Online Colloquium Speaker, University of Louisville, USA. (Invited) October 2020*
17. 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*

18. Integrative analysis of randomized clinical trial and real-world data. *JSM, virtual meeting, (Invited) August 2020*
19. 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*
20. Doubly robust inference when combining probability and nonprobability samples with high-dimensional data. *ENAR, virtual meeting. (Invited) March 2020*
21. Double score matching estimators of average treatment effects and quantile treatment effects. *Colloquium Speaker, Iowa State University, Ames, USA. (Invited) November 2019*
22. Integrative analysis of randomized clinical trial and real-world evidence studies. *Colloquium Speaker, Texas A&M University, Texas, USA. (Invited) October 2019*
23. Semiparametric Estimation of Continuous-Time Structural Failure Time Model. *JSM, Denver, Colorado, USA. (Invited) July 2019*
24. JASA, Applications and Case Studies, discussant, *JSM, Denver, Colorado, USA. (Invited) July 2019*
25. Integrative analysis of randomized clinical trial with real world evidence studies. *ICSA, Tianjin, Hebei, China. (Invited) July 2019*
26. Integrative analysis of randomized clinical trial with real world evidence studies. *Real-World Analytics, Eli Lilly and Company, Online Meeting, USA. (Invited) June 2019*
27. Causal inference with confounders missing not at random. *ICSA, Raleigh, NC, USA. (Invited) June 2019*
28. Causal inference with confounders missing not at random. *Departmental seminar speaker, Raleigh, NC, USA. April 2019*
29. Causal inference with confounders missing not at random. *ENAR, Philadelphia, PA, USA. (Invited) March 2019*
30. Causal inference with confounders missing not at random. *Colloquium Speaker, Baylor University, Waco TX, USA. (Invited) February 2019*
31. Propensity score matching and subclassification in observational studies with multi-level treatments. *The International Biometrics Society Journal Club. Webinar (Invited) December 2018*
32. Asymptotic inference of causal effects with observational studies trimmed by the estimated propensity score. *ICSA, New Brunswick, NJ, USA. (Invited) June 2018*
33. Combining multiple observational data sources to estimate causal effects. *ACIC, Pittsburgh, PA, USA. (Invited) May 2018*
34. Dynamic regime marginal structural models to survival distribution as a function of time to treatment discontinuation. *ENAR, Atlanta, GA, USA. (Invited) March 2018*
35. 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*
36. Modeling survival distribution as a function of time to treatment discontinuation. *Departmental Seminar Speaker, North Carolina State University, Raleigh, NC, USA. September 2017*
37. Modeling survival distribution as a function of time to treatment discontinuation. *Colloquium Speaker, Purdue University, West Lafayette, IN, USA. (Invited) September 2017*
38. Estimation of the cumulative incidence function under multiple dependent and independent censoring mechanisms. *JSM, Baltimore, MD, USA. (Invited) August 2017*
39. Nonparametric identification of causal effects with confounders subject to instrumental missingness. *ACIC, Chapel Hill, NC, USA. (Poster) May 2017*
40. Propensity score weighting for causal inference with multi-stage data. *ENAR, Washington, DC, USA. (Invited Poster) March 2017*
41. A note on multiple imputation of handling missing data under complex sampling. *JSM, Chicago, IL, USA. (Invited) August 2016*

42. Estimation and goodness-of-fit test of structural nested mean models. *ENAR, Austin, TX, USA. March 2016*
43. 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*
44. 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*
45. Optimal estimation of coarse structural nested mean models. *ENAR, Miami, FL, USA. March 2015*
46. 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*
47. Likelihood-based inference with missing data under missing-at-random. *JSM, Boston, MA, USA. August 2014*
48. Likelihood-based inference with missing data under missing-at-random. *ICSA-KISS symposium, Portland, OR, USA. June 2014*
49. Propensity score matching and subclassification with multivalued treatments. *ACIC, Providence, RI, USA. May 2014*
50. Fractional hot deck imputation for robust estimation under item nonresponse in survey sampling. *JSM, Montreal, Canada. August 2013*
51. Parametric fractional imputation for longitudinal data with non-ignorable missing data. *JSM, San Diego, CA, USA. August 2012*
52. Quantify uncertainty in image classification. *National Agricultural Statistics Service (NASS), Washington, DC, USA. May 2011*

Professional Experience

1. Executive committee, the 2021 DISS (Duke Industry Statistics Symposium) in Durham
2. SAMSI faculty fellow for the 2020 Causal Inference program
3. Executive committee, the 2019 ICSA symposium in Raleigh
4. Treasurer, the 2019 ICSA symposium in Raleigh
5. Organized invited sessions
    - "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*

6. Translated chapters in Book "Causal Inference for Statistics, Social and Biomedical Sciences: An Introduction" into Chinese *(2015–2017)*

Students Supervision

Current

1. Eunah Cho (PhD Advisor)
2. Sarah Reigel (PhD Advisor)
3. Chenyin Gao (PhD Advisor)
4. Yuwen Cheng (PhD Advisor)
5. Siyi Liu (PhD Advisor)
6. Yunshu Zhang (PhD Advisor)
7. Lili Wu (PhD Advisor)
8. Tanchumin Xu (PhD Co-Advisor)
9. Taekwon Hong (PhD Co-Advisor with Wenbin Lu)
10. Jianing Chu (PhD Co-Advisor with Wenbin Lu)
11. Joe Zhao (PhD Co-Advisor with Emily Hector)
12. Yichi Zhang (PhD Co-Advisor with Minh Tang)
13. Dasom Lee (PhD Co-Advisor with Sujit Ghosh)
14. Peter Norwood (PhD Committee Member)
15. Nicholas Larsen (PhD Committee Member)
16. Jesse Clifton (PhD Committee Member)
17. Can Cui (PhD Committee Member)
18. Hengrui Cai (PhD Committee Member)

Past

1. Nathan Corder (PhD Co-Advisor with Marie Davidian, 2014–2021)
2. Andrew Griffin (PhD Co-Advisor with Brian Reich, 2016–2020)
3. Lin Dong (PhD Co-Advisor with Eric Laber, 2015–2019)
4. Miao Yu (PhD Committee Member, 2017–2021)
5. Alexandra Larsen (PhD Committee Member, 2015–2019)
6. Qian Guan (PhD Committee Member, 2015–2019)
7. Eric Rose (PhD Committee Member, 2015–2019)
8. Breanna Swan (PhD Committee Member, 2016–2021)

Editorial Services

1. Co-editor, *Journal of Agricultural Biological and Environmental Statistics* Special Issue on Causal inference 2021–
2. Associate Editor, *Journal of Computational and Graphical Statistics* 2021–
3. Associate Editor, *Biometrics* 2018–2022
4. Associate Guess Editor, *Statistica Sinica* 2016–2018
5. NSF DMS panel review
6. NSERC (Natural Sciences and Engineering Research Council of Canada) grant review

7. MRC (Medical Research Council, UKRI) grant review
8. 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*

Teaching

1. ST371 Introduction to Probability and Distribution Theory
   Sections 001, 002 Fall 2016
   Sections 001, 002 Spring 2018
2. ST520 Statistical Principles of Clinical Trials
   Fall 2018
   Spring 2019

Departmental Service

1. Diversity committee (2020–2022)
2. Junior faculty representative (2019, 2020)
3. Faculty search committee (2019)
4. Beach trip committee (2019)
5. Department seminar committee (2018, 2021 Chair)
6. Qualify exam committee (2017)