Yuyao Wang

PhD candidate
University of California San Diego
La Jolla, CA 92093 U.S.A.
Email: yuw079@ucsd.edu
Website: https://wangyuyao98.github.io

Education

2019 - now *PhD*, Mathematics with a Specialization in Statistics, University of California San Diego
2019 - 2022 *MA*, Mathematics, University of California San Diego
2015 - 2019 *BS*, Mathematics, Xi'an Jiaotong University

Research Interests

Causal inference, survival analysis, missing data problems, semiparametric theory, nonparametric theory, selection bias problems. Applications to health and social sciences.

Work Experience

Jun - Aug 2024 Biostatistics internship at St. Jude Children's Research Hospital. Memphis, TN.
(Mentored by Dr. Kendrick Li)

Fellowships

2019 - 2023 UCSD Halicioglu Data Science Institute (HDSI) Graduate Prize Fellowship.

Awards

2024 Society for Causal Inference (SCI) Travel Scholarship (for 2024 ACIC).
2024 UCSD GPSA Travel Grant Award (for 2024 ACIC).
2023 Student Paper Competition Award, 2023 Lifetime Data Science Conference (LiDS).

Publications

2024 **Wang, Y.**, Ying, A., Xu, R. (2024) Doubly robust estimation under covariate-induced dependent left truncation. *Biometrika*, 111(3), 789-808. https://doi.org/10.1093/biomet/asae005
(Won the Student Paper Competition Award at 2023 Lifetime Data Science Conference)

2023 Peng, Y., **Wang, Y.**, Xu, R. (2023). Measures of explained variation under the mixture cure model for survival data. *Statistics in Medicine*, 42(3), 228-245. https://doi.org/10.1002/sim.9611

Working Papers and Preprints

2024 **Wang, Y.**, Ying, A., Xu, R. (2024) Learning treatment effects under covariate dependent left truncation and right censoring. *arXiv:2411.18879*. https://arxiv.org/abs/2411.18879

White, L. R., **Wang, Y.**, Launer, L., Seale, L., Berry, M., Flannagan, M., Montine, T.J., Xu, R. (2024+) Midlife dietary selenium and Okinawan ancestry are independently associated with less severe Alzheimer neuropathology at autopsy in Japanese-American men. *Submitted to Journal of Alzheimer's Disease.*

R packages

- `truncAIPW`: Doubly Robust Estimation under Covariate-Induced Dependent Left Truncation
- `aftR2`: R-Squared Measure under Accelerated Failure Time (AFT) Models

Presentations

2024 Learning treatment effects under covariate dependent left truncation and right censoring. *Presentation at the Online Causal Inference Seminar (OCIS).*

2024 Learning treatment effects under covariate dependent left truncation and right censoring. *Presentation in the Causal Group Seminar at Carnegie Mellon University.*

2024 Learning treatment effects under covariate dependent left truncation and right censoring. *Presentation in the Causal Inference Seminar at Boston University.*

2024 Learning treatment effects under covariate dependent left truncation and right censoring. *Presentation in topic contributed paper session at 2024 Joint Statistical Meetings (JSM).*

2024 Doubly robust estimation of treatment effects under covariate dependent left truncation and right censoring. *Poster at 2024 American Causal Inference Conference (ACIC).*

2024 Learning Treatment Effects under Covariate Dependent Left Truncation and Right Censoring. *Presentation at 2024 Southern California Applied Mathematics Symposium (COCAMS).*

2023 Doubly Robust Estimation under Covariate-induced Dependent Left Truncation. *Presentation at McGill Statistics Seminar (2023 Fall).*

2023 Doubly Robust Estimation under Covariate-induced Dependent Left Truncation. *Presentation in topic contributed paper session at 2023 Joint Statistical Meetings (JSM).*

2023 Doubly Robust Estimation under Covariate-induced Dependent Left Truncation. *Presentation at 2023 Lifetime Data Science Conference (LiDS).*

2023 Multiply Robust Estimation of Treatment Effect for Time-to-event Outcome under Dependent Left Truncation. *Poster at 2023 American Causal Inference Conference (ACIC).*

2023 Doubly Robust Estimation under Covariate-induced Dependent Left Truncation. *Poster at 2023 Public Health Research Day at UCSD.*

2022 Semiparametric Estimation for Non-randomly Truncated Data. *Poster at 2022 American Causal Inference Conference.*

2022 Semiparametric Estimation for Non-randomly Truncated Data. *Poster at 2022 Public Health Research Day at UCSD.*

Research Experience

2023 - now **Learning treatment effects under covariate dependent left truncation and right censoring**, University of California San Diego, La Jolla, CA, USA.
PhD Student | Advisor: Ronghui Xu

- Derived the efficient influence curve for a general Z-estimand under covariate-dependent left truncation.
- Developed doubly robust operators for handling covariate-dependent left truncation and right censoring that can be applied to a wide range of estimation problems, including estimating average treatment effect (ATE) and conditional average treatment effect (CATE).
- For ATE, developed model doubly robust and rate doubly robust estimators; for CATE, developed orthogonal and doubly robust learners that can achieve the oracle error rate.
- Carefully studied the asymptotics of the proposed methods, and conducted comprehensive simulations to study their finite sample performance.
- Applied the proposed approaches to analyze the effect of midlife alcohol consumption on late-life cognitive impairment, using data from the Honolulu Asia Aging Study.

2024 - now **Proximal inference for handling selection bias from dependent left truncation**, University of California San Diego, La Jolla, CA, USA.
PhD Student | Advisor: Ronghui Xu

- Developed a proximal truncation identification framework for estimating a general Z-estimand under the selection bias from dependent left truncation.

2020 - 2023 **Doubly robust estimation under covariates-induced dependent left truncation**, University of California San Diego, La Jolla, CA, USA.
PhD Student | Advisor: Ronghui Xu

- Derived the efficient influence curve (EIC) for the expectation of an arbitrarily transformed survival time.

- Constructed EIC-based estimators that are shown to have favorable properties, including model double robustness, rate double robustness, and semiparametric efficiency.
- Provided technical conditions for the asymptotic properties that appear to not have been carefully examined in the literature for time-to-event data.
- The work represents the first attempt to construct doubly robust estimators in the presence of left truncation, which is a common source of selection bias for time-to-event data that does not fall under the established framework of coarsened data where doubly robust approaches were developed.
- Conducted comprehensive simulations to study the finite sample performance of the proposed estimators.
- Applied our estimator to analyze data on central nervous system (CNS) lymphoma and data on cognitive impairment from Honolulu Asia Aging study.

2024 **Selection bias adjustment for time-to-event analysis of prospective childhood cancer survivorship studies**, St. Jude Children's Research Hospital, Memphis, TN, USA.
Biostatistics intern | Supervisor: Kendrick Li

- Developed inverse probability weighting approaches for handling the selection bias from covariate-dependent left truncation due to delayed entry in semi-competing risks settings.
- Conducted comprehensive simulations to compare the finite sample performance of the proposed approaches to the conventional approaches.
- Applied the proposed methods to analyze the cardiotoxicity of childhood cancer treatments, such as anthracyclines and radiation, under Cox proportional hazards models, using data from SJLIFE.
- Derived the efficient influence curve for a general Z-estimand under semi-competing risks settings with covariate dependent left truncation.

2023 - 2024 **Investigation of factors that protect against the development of Alzheimer's Disease neuropathological changes**, University of California San Diego, La Jolla, CA, USA.
Research Assistant | Supervisors: Ronghui Xu, Lon White

- Data visualization and data analysis for Alzheimer brain autopsy data from Honolulu Asia Aging Study.
- Studied risk factors (FEV1, dietary selenium, dietary maltose) for Alzheimer brain pathology at death.
- Applied inverse probability weighting approaches for handling the selection bias from double truncation that are involved in the data.

2023 **Assessing safety of COVID-19 vaccine exposure during pregnancy**, UC San Diego based Organization of Teratology Information Services (OTIS) research center, San Diego, CA, USA.
Statistical Consultant | Supervisors: Ronghui Xu, Jennifer (Yunjun) Luo

- Contributed to develop statistical analysis approaches for assessing safety of medications during pregnancy; particularly in addressing challenges in analyzing the effect of different timings of exposure in the presence of the selection bias due to delayed entry (left truncation).

2020 - 2022 **Measures of Explained Variation under the Mixture Cure Model**, University of California San Diego, La Jolla, CA, USA.
PhD Student | Advisor: Ronghui Xu

- Proposed two approaches to define explained variation under mixture cure models: based on the Kullback-Leibler information gain, and based on residual sum of squares.
- Studied properties of the proposed measures both analytically and with simulation studies
- Applied the proposed approaches to analyze data on bone marrow transplant and the SEER-medicare data.

2020 - 2021 **Covariate balancing weights for estimating average treatment effect with censored time-to-event data**, University of California San Diego, La Jolla, CA, USA.
Student researcher | Supervisor: Jelena Bradic

- Reviewed the literature on covariate balancing methods and investigated the possibility of developing covariate balancing weights to estimate average treatment effect for censored time-to-event data.
- Developed method that use balancing weights for confounding and use inverse probability of censoring weighting to handle informative right censoring.

Professional Service

2024 Reviewer for Biometrics.

2024 Organizer of the invited session "Recent Advances in Causal Inference to Address Complexities in Real-world Data" at 2025 Joint Statistical Meetings (JSM).

2024 Member of the Scientific Organizing committee for 2025 International Conference on Health Policy Statistics (ICHPS).

2024 Organizer of the topic-contributed paper session "Causal Inference with Complex Missing Patterns" at 2024 Joint Statistical Meetings (JSM).

2024 UCSD Association of Women in Math (AWM) 2024 - 2025 Colloquium coordinator.

Teaching Experience

University of California San Diego, La Jolla, CA, USA.

2024 MATH 181B: Introduction to Mathematical Statistics, TA, Fall.
MATH 284: Lifetime Data Analysis, TA, Spring.
FMPH 291: Special Topics/Public Health - Applied Survival Analysis, TA, Spring.
DSC 180B: Data Science Project II, TA, Winter.

2023 DSC 180A: Data Science Project I, TA, Fall.
High School Math Program (Probability and Statistics, advanced track), mentor, Summer.
MATH 284: Lifetime Data Analysis, TA, Spring.
MATH 181B: Introduction to Mathematical Statistics, TA, Spring.
DSC 180B: Data Science Project II, TA, Winter.

2022	DSC 180A: Data Science Project I, TA, Fall.
	MATH 181A: Introduction to Mathematical Statistics I, TA, Winter and Spring.
2021	MATH 10A: Calculus I, TA, Fall.
	MATH 185: Introduction to Computational Statistics, TA, Spring.
	MATH 189: Data Analysis and Inference, TA, Winter.
2020	MATH 11: Calculus-Based Probability and Statistics, TA, Fall.
	MATH 189: Data Analysis and Inference, TA, Winter and Spring.
2019	MATH 10B: Calculus II, TA, Fall.

Outreach

2024	Mentor in the Math Graduate Mentorship Program at UC San Diego
2023	Presentation at the UCSD Halicioglu Data Science Institute research review event for industry partners
2023	Poster presentation at the UCSD Halicioglu Data Science Institute Open House for prospective PhD students
2022	Moderator for the Prospective International Graduate Student Panel in math department at UC San Diego
2022	Mentor in AWM undergrad mentorship program at UC San Diego
2020 - 2021	Mentor in the Math Graduate Mentorship Program at UC San Diego

Last updated: December 2, 2024 • Typeset in XeLaTeX