CHENYIN GAO

cgao@hsph.harvard.edu | (919) 396-3613 | Boston, MA | Homepage

WORK EXPERIENCE

Harvard University, USA 2024 – present
Postdoctoral Research Fellow in Biostatistics
Adviser: Dr. Rui Duan, rduan@hsph.harvard.edu

Eli Lilly & Company 2022 – 2025
Academic Research Intern

Duke University 2021
Summer Research Intern

RESEARCH INTEREST

My research focuses on developing innovative statistical methods to make causal inferences for complex quasi-experiments and observational studies, including long-term causal effect estimation, sequential treatment design, and large-scale data integration. In practice, the developed methods are applicable in various fields, such as business, clinical trials, environmental health, and cancer research, to identify effective personalized interventions.

EDUCATION

North Carolina State University, USA Aug. 2019 – July 2024
- Ph.D. in Statistics. GPA: 4.0/4.0
- Thesis: Advanced Statistical Methods for Data Integration and Tensor Completion in Causal Inference [link]
- Advisor: Dr. Shu Yang, syang24@ncsu.edu

Sun Yat-sen University, P.R. China Aug. 2015 – June 2019
- B.Sc. in Statistics. GPA: 3.8/4.0. Minor in Finance

AWARDS AND HONORS

- Student Paper Award, LiDS, 2025
- Student Paper Award, ASA, 2025
- MBB Interdisciplinary Mind Grant, Harvard, 2024
- Student Paper Award, ICSA, 2024
- Paige Plagge Graduate Award for Citizenship, NCSU, 2024
- Best Poster Award, DISS, 2024
- Student and Early-Career Travel Award, JSM, 2023
- China National Scholarship, China, 2018 (Awarded for outstanding full-time undergraduates)

PUBLICATIONS

* Correspondence author

1. **C. Gao**, L. Han*, and P. Gilbert (2025). Bridging Fairness and Efficiency in Conformal Inference: A Surrogate-Assisted Group-Clustered Approach, *International Conference on Machine Learning (ICML)*
2. **C. Gao**, S. Yang*, M. Shan, W. Ye, I. Lipkovich, D. Faries (2025). Doubly protected estimation for survival outcomes utilizing external controls for randomized clinical trials, *International Conference on Machine Learning (ICML)*
3. **C. Gao,** X. Zhang, S. Yang* (2025) Omnibus doubly robust sensitivity analysis of externally controlled trials with intercurrent events, *Biometrics*
4. **C. Gao**, S. Yang*, M. Shan, W. Ye, I. Lipkovich, D. Faries (2025). Improving randomized controlled trial analysis via data-adaptive borrowing, *Biometrika*
5. D. Faries*****, **C. Gao**, X. Zhang, C. Hazlett, J. Stamey, S. Yang, P. Ding, M. Shan, K. Sheffield, N. Dreyer (2024) Real Effect or Bias? Best Practices for Evaluating the Robustness of Real-World Evidence through Quantitative Sensitivity Analysis for Unmeasured Confounding, *Pharmaceutical Statistics*
6. L. Wu, **C. Gao**, 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 (JRSSC)*
7. **C. Gao***, Z. Zhang, and S. Yang (2024). Causal Customer Churn Analysis with Low-rank Tensor Block Hazard Model, *International Conference on Machine Learning (ICML)*
8. **C. Gao***, S. Yang, and A. Zhang (2024). Enhancing convolutional neural network generalizability via low-rank weight approximation, *IET Image Processing*
9. 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 (JBS)*

10. **C. Gao*** and S. Yang (2023). Pretest estimation in combining probability and non-probability samples, *Electronic Journal of Statistics (EJS)*
11. **C. Gao**, S. Yang*****, and J. K. Kim (2023). Soft calibration for correcting selection bias under mixed-effects models, *Biometrika*
12. 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 (JRSSB)*
13. **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 (JRSSA)*
14. Q. Xie*****, T. Du, M. Zhao, **C. Gao**, Q. Lyu, L. Suo, Y. Kuang (2021). Advanced trophectoderm quality increases the risk of a large for gestational age baby in single frozen-thawed blastocyst transfer cycles. Human Reproduction
15. Y. Deng, **C. Gao*** (2022). Where does the risk lie? Systemic risk and tail risk networks in the Chinese financial market. Pacific Economic Review

PREPRINTS

16. **C. Gao**, A. Zhang, and S. Yang***** (202X). Causal inference on sequential treatments via tensor completion, *in revision*
17. I. Lipkovich*****, Z. Kadziola, **C. Gao**, D. Wang, D. Faries (202X) Evaluation of machine learning approaches for estimating optimal individualized treatment regimens for time-to-event outcomes in observational studies
18. **C. Gao**, P. B Gilbert, and L. Han (202X) On the role of surrogates in conformal inference of individual causal effects
19. **C. Gao**, J. D. Tubbs, Y. Han, M. Guo, S. Li, E. Ma, D. Luo, J. W. Smoller, P. H. Lee, and R. Duan (202X) Unsupervised ensemble learning for efficient integration of pre-trained polygenic risk scores

PRESENTATION

- Doubly protected estimation for survival outcomes utilizing external controls for randomized clinical trials. *JSM (2025), Nashville*, TN
- Doubly protected estimation for survival outcomes utilizing external controls for randomized clinical trials. *LiDS (2025), New York City*, NY
- Causal Inference on Sequential Treatments via Tensor Completion. *JSM* (*2024*), Portland, OR
- Improving randomized controlled trial analysis via data-adaptive borrowing. *ICSA (2024)*, Nashville, TN (Invited)
- Causal Inference on Sequential Treatments via Tensor Completion. *The New England Statistics Symposium* (*2024*), University of Connecticut (Invited)
- Integrating Randomized Trial Data with External Controls: A Semiparametric Approach with Selective Borrowing. *JSM (2023)*, Toronto, ON (Invited)
- Semi-parametric efficient integrative estimator with historical controls. *ENAR (2023)*, Nashville, TN
- Causal Inference on Sequential Treatments via Tensor Completion. *(2023)*, Duke University, Durham, NC (Invited)
- Pretest estimation in combining probability and non-probability samples. *JSM (2022)*, Washington, DC
- CNN-based Single Cryo-EM Images Unsupervised Denoisers. Impact Talk presented at: *17th Annual Duke Center for AIDS Research Virtual Fall Scientific Retreat (2021),* Durham, NC (virtual)

SKILLS

- **Software:** R, Python, PyTorch, SQL, SAS
- **R/Python package:**
  - ElasticIntegrative implements elastic analyses for the heterogenous treatment effects combining trials and real-world data.
  - SelectiveIntegrative implements dynamic borrowing framework to incorporate information from other external-control (EC) datasets with the gold-standard randomized trials.
  - TensorBlockHazard implements the tensor factor model with clustering structure to analyze customer churn.
  - SurrConformalDR implements the surrogate-assisted doubly robust conformal inference for individualized treatment effect evaluation.
- **Language:** Chinese (native), English
- **Others:** CFA Level I (link)

ACTIVITIES AND SERVICES

- Invited Chair for *JSM 2024*
- Top reviewer for *AISTATS (2022, 2023, 2024)*
- Co-organizer for *BIRS (Banff International Research Station) 5-day workshop, May 22–27, 2022* "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