CHENYIN GAO

cgao@hsph.harvard.edu | (919) 396-3613 | Raleigh, NC | Homepage

WORK/RESEARCH EXPERIENCE

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

Eli Lilly & Company 2022 – present
Academic Research Intern

Duke University 2021
HIV/AIDS Research Intern

EDUCATION

North Carolina State University, USA Aug. 2019 – July 2024
- Ph.D. in Statistics. GPA: 4.0/4.0
- Research interests: causal inference, data integration, precision medicine, missing data and tensor analysis
- 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, ICSA, 2024
- Paige Plagge Graduate Award for Citizenship, NCSU, 2024
- Best Poster Award, DISS, 2024
- Student and Early-Career Travel Award, JSM, 2023
- Mu Sigma Rho, National Statistics Honor Society, NCSU, 2021 - present
- Excellent Undergraduate Dissertation, SYSU, 2019
- China National Scholarship, China, 2018 (Awarded for outstanding full-time undergraduates)
- 1st Merit Scholarship, School of Mathematics, SYSU, 2018 (2/72)
- 1st Prize, China Undergraduate Mathematical Contest in Modeling, CSIAM, 2018

PUBLICATIONS/PREPRINTS

* Correspondence author

1. L. Wu, **C. Gao**, S. Yang*, B. J. Reich, and A. Rappold. Estimating spatially varying health effects in app-based citizen science research (2024), *Journal of the Royal Statistical Society: Series C*, accepted, DOI: 10.1093/jrsssc/qlae034
2. **C. Gao***, Z. Zhang, and S. Yang (2024). Causal Customer Churn Analysis with Low-rank Tensor Block Hazard Model, *International Conference on Machine Learning.*
3. **C. Gao***, S. Yang, and A. Zhang (2024). Enhancing convolutional neural network generalizability via low-rank weight approximation, *IET Image Processing*, accepted.
4. 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: 10.1080/10543406.2024.2330216
5. **C. Gao*** and S. Yang (2023). Pretest estimation in combining probability and non-probability samples, *Electronic Journal of Statistics*, DOI: 10.1214/23-ejs2137
6. **C. Gao**, S. Yang*, and J. K. Kim (2023). Soft calibration for correcting selection bias under mixed-effects models, *Biometrika*, DOI: 10.1093/biomet/asad016
7. 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*, DOI: 10.1093/jrsssb/qkad017
8. **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*, DOI: 10.1111/rssa.12841
9. **C. Gao**, A. Zhang, and S. Yang* (202X). Causal inference on sequential treatments via tensor completion, *submitted*
10. **C. Gao**, S. Yang*, M. Shan, W. Ye, I. Lipkovich, D. Faries (202X) Integrating Randomized Placebo-Controlled Trial Data with External Controls: A Semiparametric Approach with Selective Borrowing, *submitted*
11. **C. Gao,** X. Zhang, S. Yang* (202X) Omnibus sensitivity analysis of externally controlled trials with intercurrent events, *submitted*
12. 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, *submitted*

13. D. Faries*****, **C. Gao**, X. Zhang, C. Hazlett, J. Stamey, S. Yang, P. Ding, M. Shan, K. Sheffield, N. Dreyer (202X) Real Effect or Bias? Best Practices for Evaluating the Robustness of Real-World Evidence through Quantitative Sensitivity Analysis for Unmeasured Confounding, *submitted*
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 36: 2111-2120
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, DOI: 10.1111/1468-0106.12417

PRESENTATION

- C. Gao, A. Zhang, S. Yang, Causal Inference on Sequential Treatments via Tensor Completion. *JSM* (*2024*), Portland, OR
- C. Gao, S. Yang, M. Shan, W. Ye, I. Lipkovich, and D. Faries, Improving randomized controlled trial analysis via data-adaptive borrowing. *ICSA (2024)*, Nashville, TN (Invited)
- C. Gao, A. Zhang, S. Yang, Causal Inference on Sequential Treatments via Tensor Completion. *The New England Statistics Symposium* (*2024*), University of Connecticut (Invited)
- 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. *JSM (2023)*, Toronto, ON (Invited)
- C. Gao, S. Yang, Semi-parametric efficient integrative estimator borrowing historical controls with penalized bias. *ENAR (2023)*, Nashville, TN
- C. Gao, A. Zhang, S. Yang, Causal Inference on Sequential Treatments via Tensor Completion. *(2023)*, Duke University, Durham, NC (Invited)
- C. Gao, S. Yang, Pretest estimation in combining probability and non-probability samples. *JSM (2022)*, Washington, DC
- C. Gao, P. Acharya, A. Zhang, 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
- **Language:** Chinese (native), English
- **Others:** CFA Level I (link)

ACTIVITIES AND SERVICES

- Invited Chair for *JSM 2024*
- Top reviewer for *AISTATS (2022, 2023)*
- 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