Ruoyu Wang

E-mail: ruoyuwang@hsph.harvard.edu | Phone Number: +1 6172596271
Address: Department of Biostatistics, Harvard University
677 Huntington Avenue, Boston, MA 02115

Education

Ph.D in Probability and Mathematical Statistics — Sept, 2017 - May 23, 2022
Academy of Mathematics and Systems Science (AMSS),Chinese Academy of Sciences — *Beijing, China*

Supervisor: Qihua Wang

Bachelor in Statistics; Top 2% — Sept, 2013 – Jun, 2017
Nankai University — *Tianjin, China*

Work Experience

Postdoctoral Fellow — Sept, 2022 – present
Department of Biostatistics, Harvard University — *Boston, United States*

Supervisor: Xihong Lin

Research Interests

- Data Fusion
- Causal Identification and Inference
- Mendelian Randomization
- Large-scale Data Analysis
- Two-phase Sampling
- Statistical Learning Theory

Peer-Reviewed Publications

- **Wang, R.**, Wang Q.*, and Miao, W. (2023), A robust fusion-extraction procedure with summary statistics in the presence of biased sources. *Biometrika*, 110, 1023–1040.
- **Wang, R.**, Su, M., and Wang, Q.* (2023), Distributed nonparametric imputation for missing response problems with massive data. *Journal of Machine Learning Research*, 68, 1–52.
- **Wang, R.**, and Wang, Q.* (2021), Determination and estimation of optimal quarantine duration for infectious diseases with application to data analysis of COVID-19. *Biometrics*, 78, 691–700.
- **Wang, R.**, Wang, Q.*, Miao, W., and Zhou, X. (2024), Sharp bounds for variance of treatment effect estimators in the finite population in the presence of covariates. *Statistica Sinica*, 34, 999–1021.
- **Wang, R.**[1], Yi, M.[1], Chen, Z., and Zhu, S. (2022), Out-of-distribution generalization with causal invariant transformations. *IEEE Conference on Computer Vision and Pattern Recognition*, 375–385.
- Yi, M.[1], **Wang, R.**[1], and Ma, Z. (2022), Characterization of excess risk for locally strongly convex population risk. Advances in Neural Information Processing Systems 36.
- Yi, M., **Wang, R.**, Sun, J., Li, Z., and Ma, Z. (2023), Breaking correlation shift via conditional invariant regularizer. In Proceedings of the 11th International Conference on Learning Representations.

- Yang, H., Liu, Z., **Wang, R.**, Lai, E., Schwartz, J., Baccarelli, A., Huang, Y. and Lin, X.* (2025), Causal Mediation Analysis for Integrating Exposure, Genomic, and Phenotype Data. *Annual Review of Statistics and Its Application*, 12, 337–360.

- Su, M.[1], **Wang, R.**[1], and Wang, Q.* (2022), A two-stage optimal subsampling estimation for missing data problems with large-scale data. *Computational Statistics and Data Analysis*, 173.

- Wang, Q., Su, M.*, and **Wang, R.** (2021), A beyond multiple robust approach for missing response problem. *Computational Statistics and Data Analysis*, 155.

- Miao, W. *, Li, W., Hu, W., **Wang, R.**, and Geng, Z. (2021), Invited commentary: Estimation and bounds under data fusion. *American Journal of Epidemiology*, 191, 674–678.

Paper under Invited Revision

- **Wang, R.**, Zhang, H., and Lin X.* (2025+), Debiased Estimating Equation Method for Versatile and Efficient Mendelian Randomization Using Large Numbers of Correlated Weak and Invalid Instruments. Revision invited by *Journal of the American Statistical Association: T&M*. arXiv:2408.05386.

- Hu, W.[1], **Wang, R.**[1], Li, W.*, and Miao, W.* (2025+), Semiparametric Efficient Fusion of Individual Data and Summary Statistics. Revision invited by *Journal of the American Statistical Association: T&M*. arXiv:2210.00200.

- Su, M. and **Wang, R.*** (2025+), Subsampled One-Step Estimation for Fast Statistical Inference. Revision invited by *Scandinavian Journal of Statistics*. arXiv:2407.13446.

- **Wang, R.**, Wang Q.*, and Miao, W. (2025+), A maximin optimal approach for sampling designs in two-phase studies. Revision invited by *Statistica Sinica*. arXiv:2312.10596.

Paper under Review

- **Wang, R.** and Miao, W.* (2025+), Extreme-based causal effect learning with endogenous exposures and a light-tailed error. Under review. arXiv:2408.06211.

- Su, M. and **Wang, R.*** (2025+), Moment-assisted Subsampling based Maximum Likelihood Estimator. Under review. arXiv:2309.09872.

- Yi, M., Matabuena, M., **Wang, R.*** (2025+), Denoising data with measurement error using a reproducing kernel-based diffusion model. Under review. arXiv:2501.00212.

- Su, M. and **Wang, R.*** (2025+), Moment-assisted subsampling method for Cox proportional hazards model with large-scale data. Under review. arXiv:2501.06924.

- **Wang, R.** and Lin X.* (2025+), Divide-and-shrink: An efficient and heterogeneity-agnostic approach for transfer estimation using summary statistics. Under review.

- Yang, H.[1], **Wang, R.**[1], Lin, Y., and Lin, X.* (2025+), Tail Likelihood Ratio Method for Large-Scale Causal Mediation Testing in Epigenome-Wide Studies. Under review.

- Zhang, P., **Wang, R.**, and Miao, W.* (2025+), Causal attribution with confidence. Under review. arXiv:2504.08294.

[1] Equal contribution.
* Corresponding author.

Reviewer

Journal of the American Statistical Association (JASA); *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*; *Biometrics*; *Journal of Computational and Graphical Statistics*; *Statistics in Medicine*; *IEEE Conference on Computer Vision and Pattern Recognition*.

Awards

• 2021.9	AMSS	**President Scholarship, Grand Prize**
• 2021.5	AMSS	**Merit Student**
• 2020.5	AMSS	**Merit Student**
• 2016.12	Nankai University	**First Prize Scholarship**
• 2015.12	Nankai University	**Merit Student**

Visit

- Department of Statistics, Rutgers University. March, 2025.

Oral Presentations

- Characterization of excess risk for locally strongly convex population risk. Chinese Association for Applied Statistics (CAAS), High Dimensional Statistics Symposium, July, 2021.
- Sharp bounds for variance of the treatment effect estimators in finite population in the presence of covariates. The 2021 International Workshop on Statistical Theory and Related Fields (STARF 2021), December, 2021.
- Debiased estimating equation method for summary statistics-based Mendelian randomization. The 1st Joint Conference on Statistics and Data Science in China, July, 2023.
- DEEM: A Flexible and Efficient Method for Summary Statistics-based Mendelian Randomization. ENAR 2024 Spring Meeting, March, 2024.
- Divide-and-shrink: a heterogeneity-agnostic approach for safe data integration. **Invited talk**, IMS-China 2024, July, 2024.
- Extreme-based causal effect learning with endogenous exposures and a light-tailed error. **Invited talk**, 2024 International Conference on Frontiers of Data Science, July, 2024.
- Debiased Estimating Equation Method for Versatile and Efficient Mendelian Randomization Using Large Numbers of Correlated Weak and Invalid Instruments. The 2nd Joint Conference on Statistics and Data Science in China, July, 2024.
- Debiased Estimating Equation Method for Versatile and Efficient Mendelian Randomization. 2024 Joint Statistical Meetings (JSM), August, 2024.
- Extreme-based causal effect learning with endogenous exposures and a light-tailed error. Seminar of Center for Causal Inference, University of Pennsylvania, March, 2025.
- Divide-and-shrink: An efficient and heterogeneity-agnostic approach for transfer estimation using summary statistics. **Invited talk**, Diabetes StatClin Meeting, April, 2025.

Service

- Session Chair for Joint Statistical Meeting, Portland, OR, 2024.