Yimeng Shang

(+1) 646-704-5390 | yqs5519@psu.edu | ys3298.github.io

Education

Pennsylvania State University — *Hershey, PA*
Ph.D. in Biostatistics (GPA: 4.0/4.0) — *2021.08 - 2025.05*
Advisor: Dr. Lan Kong

Columbia University — *New York, NY*
M.S. in Biostatistics (GPA:4.0/4.0) — *2019.08 - 2021.06*

East China Normal University — *Shanghai, China*
B.S. in Mathematics — *2015.09 - 2019.06*

University of California, Berkeley — *Berkeley, CA*
International Study Program — *2017.08 - 2018.05*

Work Experience

Merck & Co., Inc. — *Upper Gwynedd, PA*
Biostatistics Intern, BARDS — *2024.06 - 2024.08*
- Examined various indirect treatment comparison (ITC) methods, including Bucher method, Simulated Treatment Comparison (STC), and Matching-Adjusted Indirect Comparison (MAIC), for longitudinal outcomes across different trial populations through simulation studies.
- Proposed Arm-based MAIC to preserve the balance between arms in the reweighted population, which showed more accurate and precise estimation, better controlled Type I error and greater statistical power, compared to conventional MAIC.
- Applied and compared different ITC methods in real vaccine trials.

Cytel, Inc. — *Remote, US*
Strategic Consulting Intern — *2022.06 - 2022.08*
- Proposed predictive variable/biomarker selection algorithm controlling multiple comparisons via knockoff filters.
- Developed Shiny App to implement the proposed algorithm.
- Supported early phase dose escalation and cohort expansion simulation and wrote statistical analysis plan for FDA submission.

Eli Lilly & Co., Inc. — *Shanghai, China*
Data Science & Solution Intern — *2018.09 - 2019.06*
- Supported data management work in clinical trials including data cleaning and missing data query under the supervision of the China DSS team.
- Constructed quantitative analysis and developed an automatic web page with Shiny APP for reproducible monthly analysis to improve efficiency.

Research Experience

Causal Effect Estimation under Misclassified Treatment Assignment: a Latent Variable Approach via Outcome Modeling — *Penn State University*
Research assistant, supervised by Dr. Lan Kong — *2024.03 - Now*
- Proposed a latent variable approach for causal effect estimation with potential misclassification of treatment assignments by decomposing the likelihood function into three components: propensity score model, measurement error model, outcome model, and estimated model parameters using expectation–maximization (EM) algorithm.
- Incorporated validation data and machine learning approach to enhance the measurement error modeling and doubly-robust estimation for propensity score model and outcome model.
- Demonstrated the superiority of the proposed framework in reducing the bias caused by misclassification, especially when utilizing a machine learning algorithm for the measurement error model by simulation studies.

Robust Propensity Score Estimation via Loss Function Calibration — *Penn State University*
Research assistant, supervised by Dr. Lan Kong — *2023.03 - 2024.03*
- Proposed a calibration-based method to estimate the propensity score (PS) by incorporating a covariate imbalance penalty into the loss function of predictive PS models, such as logistic regression or neural networks.
- Demonstrated the robustness of the proposed method to PS model misspecification by showing a significant reduction in bias and RMSE through comprehensive simulation studies.

Robust Propensity Score Estimation via Loss Function Calibration for High-dimensional Real World Data (RWD)

Penn State University

Research assistant, supervised by Dr. Lan Kong *2023.06 - 2024.08*

- Extended the *loss function calibration* method to high-dimensional setting by incorporating outcome-adaptive variable selection for propensity score model.
- Extracted cohorts with high-dimensional covariates to emulate clinical trials data using real-world MarketScan Claims database.
- Conducted Plasmode simulation with extracted RWD to evaluate the proposed high-dimensional method.
- Demonstrated that the proposed method surpasses others in providing unbiased causal effect estimation.

Non-Parametric Analysis of Transient Data: a Pseudo-Competing Event Approach

Penn State University

Research assistant, supervised by Dr. Shouhao Zhou *2022.10 - 2023.12*

- Proposed a novel non-parametric approach to enhance estimation and hypothesis testing for transient data by conceptualizing state transitions as pseudo-competing events and reframing the analysis as a competing events problem.
- Calibrated the cumulative incidence estimation by inverse probability weighting to eliminate systematic bias from the pseudo-competing transition risks.
- Demonstrated unbiased estimation with accurate type I error control and robust statistical power by simulation studies.
- Developed Shiny App for application.

Statistical Analysis of High Dimensional Metabolomics Data in Autism Spectrum Disorder (ASD)

Columbia University

Research assistant, supervised by Dr. Xiaoyu Che *2020.06-2020.10*

- Constructed logistic regression and Cox hazard model to estimate the effect size for each biomarker; adjusted for multiple comparisons using Hochberg step-up method; conducted power analysis to compare the models and did sensitivity analysis by adjusting for potential confounding variables and testing interaction terms.
- Applied Bayesian generalized linear models to calculate credible intervals and select analytes with large Bayesian factors.
- Implemented Adaptive LASSO, Random Forest, and XGBoosting algorithms as feature selection methods with Bootstrap for a robust predictive model.

Publications

Shang Y, Chiu Y, Kong L. "Robust Propensity Score Estimation via Loss Function Calibration". Statistical Methods in Medical Research (Under Revision). 2024

Shang Y, Ning J, Minagawa K, Zhou S. "Non-Parametric Analysis of Transient Data: a Pseudo-Competing Event Approach". Statistical Methods in Medical Research (Under Review). 2024

Che, X., Roy, A., Bresnahan, M., Mjaaland, S., Reichborn-Kjennerud, T., Magnus, P., Stoltenberg, C., **Shang, Y.**, Zhang, K., Susser, E., Fiehn, O., & Lipkin, W. I. "Metabolomic analysis of maternal mid-gestation plasma and cord blood in autism spectrum disorders". Molecular psychiatry, 2023; 28(6):2355–2369. doi:10.1038/s41380-023-02051-w

Endres KM, Kierys K, **Shang Y**, et al. A Multicenter Retrospective Evaluation of Specialized Laboratory Investigations in the Workup of Pediatric Patients With New-Onset Supraventricular Tachycardia. J Emerg Nurs. 2022;48(6):678-687. doi:10.1016/j.jen.2022.07.002

Abdalla M, Chiuzan C, **Shang Y**, et al. Factors Associated with Insomnia Symptoms in a Longitudinal Study among New York City Healthcare Workers during the COVID-19 Pandemic. Int J Environ Res Public Health. 2021;18(17):8970. doi:10.3390/ijerph18178970

Shechter A, Chiuzan C, **Shang Y**, et al. Prevalence, Incidence, and Factors Associated with Posttraumatic Stress at Three-Month Follow-Up among New York City Healthcare Workers after the First Wave of the COVID-19 Pandemic. Int J Environ Res Public Health. 2021;19(1):262. doi:10.3390/ijerph19010262

Skills

Statistics Causal inference, Clinical trials, Variable selection, Survival analysis, Measurement error, Machine learning, Medical collaborative data analysis, Real world data/evidence (Claims/EHR data)

Programming R (base R, Tidyverse, ggplot, RShiny, ggsurvfit, parallel computing), Python (Pytorch), SAS, STATA, Bash, Linux

Awards

Spring 2024 Travel Award Penn State College of Medicine