

Yimeng Shang

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

Education

Pennsylvania State University Ph.D. in Biostatistics (GPA: 4.0/4.0); Advisor: Dr. Lan Kong	Hershey, PA 2021.08 - 2025.03 (Expected)
Columbia University M.S. in Biostatistics (GPA:4.0/4.0)	New York, NY 08/2019 - 06/2021
East China Normal University B.S. in Mathematics (GPA:3.5/4.0)	Shanghai, China 09/2015 - 06/2019

Industry Experience

Merck & Co., Inc. Biostatistics Intern, BARDS (HTA, Vaccine)	Upper Gwynedd, PA 06/2024 - 08/2024
<ul style="list-style-type: none">Examined the impact of varying covariate overlap across diverse trial populations on indirect treatment comparison (ITC) methods, including the Bucher method, Simulated Treatment Comparison (STC), and Matching-Adjusted Indirect Comparison (MAIC), to analyze longitudinal outcomes using comprehensive 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.Implemented and evaluated various ITC methods to compare the effect of pneumococcal vaccines V114 and PCV20 using clinical trial data (SDTM, ADaM).	
Cytel, Inc. Strategic Consulting / Biostatistics Intern	Boston, MA 06/2022 - 08/2022
<ul style="list-style-type: none">Proposed a predictive variable/biomarker selection algorithm for subgroup identification using knockoff filters to control for multiple comparisons and built an interactive Shiny app to facilitate the use of the proposed algorithm.Supported dose escalation and cohort expansion simulations and prepared the statistical analysis plan for FDA submission.	
Eli Lilly & Co., Inc. Data Science & Solution Intern	Shanghai, China 09/2018 - 06/2019
<ul style="list-style-type: none">Assisted with data management in clinical trials, including data cleaning and addressing missing data queries.Conducted quantitative analysis and developed an automated Shiny app for reproducible monthly analysis to enhance efficiency.	

Skills

Statistics	Causal inference, (Clinical) trials design and emulation, Survival analysis and Competing events, Real-world data/evidence, High dimensional data analysis, Measurement error, Machine learning, Bayesian analysis, statistical computations and simulations using high-performance clusters.
Programming	R (base R, Tidyverse, ggplot, RShiny, Rmarkdown, ggsurvfit, parallel computing), Python (Pytorch), SAS, Bash, Linux.
Data Experience	Real-world data (Electronic health records, Claims databases), Clinical trials data (CDISC, SDTM, ADaM), Observational data (UK Biobank), Omics data

Awards

2025 Student Paper Award (Honorable Mention)	ASA Risk Analysis Section
Travel Award (2024, 2025)	Penn State College of Medicine
Scored the highest in Ph.D. qualifying exam	Penn State College of Medicine

Academic Presentations

ENAR 2024 (Poster presentation)	Robust Propensity Score Estimation via Loss Function Calibration
ENAR 2025 (Oral presentation)	A Latent Variable Approach for Causal Effect Estimation under Misclassified Treatment
JSM 2025 (Invited oral presentation)	Non-Parametric Analysis of Transient Data: a Pseudo-Competing Event Approach

Publications

- **Shang Y**, Chiu Y, Kong L. “Robust Propensity Score Estimation via Loss Function Calibration”. *Statistical Methods in Medical Research*, in press. 2024
- **Shang Y**, Ning J, Minagawa K, Zhou S. “Non-Parametric Analysis of Transient Data: a Pseudo-Competing Event Approach”. *Statistics in Medicine* (Under Review). 2025. **Won 2025 ASA Student Paper Award (Honorable Mention), Risk Analysis Section.**
- **Shang Y**, Chiu Y, Kong L. “A Latent Variable Approach for Causal Effect Estimation under Misclassified Treatment Assignment”. 2025. (Plan to submit to *Statistics in Medicine*).
- **Shang Y**, Kim Y, Mt-Isa S, Li J. “Assessing the performance of indirect treatment comparison methods for longitudinal outcomes”. 2025. (In preparation).
- **Shang Y**, Ning J, Minagawa K, Zhou S. “GRAT: Graphical Representation and Analysis of Transient data”. 2025. (In preparation).
- Kurapati, S.S., Du, A., Bowie, E.M. Scott, I.U., **Shang, Y.**, Kong, L., Das, A.V. “Global Policy Lens: Associations between Inception of National Vision Health Policy Programs, Country Indicators, & Prevalence of Blindness”. *American Society of Cataract and Refractive Surgery Annual Meeting*. 2025.
- Zhang R, **Shang Y**, Cioccio J... “Sensitivity and specificity of chimerism tests in predicting leukemia relapse using increasing mixed chimerism”. *The Journal of Molecular Diagnostics*, 2024
- Slobodanka P, **Shang Y**, Alexandors V ... “C-reactive protein improves the ability to detect hypertension and insulin resistance in mild-to-moderate obstructive sleep apnea: age effect”. *Journal of Sleep Research*, 2024
- Vgontzas A, **Shang Y**, He F... 0392 “Insomnia with Short Sleep Duration Is Associated with Heart Disease and Stroke: Evidence from the UK Biobank Cohort”. *Sleep*. 2024
- Che X,..., **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
- Endres KM, Kierys K, **Shang Y**... A Multicenter Retrospective Evaluation of Specialized Laboratory Investigations in the Workup of Pediatric Patients With New-Onset Supraventricular Tachycardia. *J Emerg Nurs*. 2022
- Abdalla M, Chiuzan C, **Shang Y**... 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
- 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

Research Projects

Estimating Per-Protocol Effects in Randomized Controlled Trials with Survival Outcomes and Competing Events: Addressing Non-Adherence

Penn State University

Supervised by Dr. Yu-Han Chiu

08/2024 - Present

- Conducted statistical analysis for the COSMOS trials to estimate the per-protocol and intent-to-treat effects of cocoa flavanol supplementation on preventing cardiovascular disease (CVD) events, accounting for non-CVD deaths as competing events.
- Developed and applied an inverse probability weighting (IPW) estimator and g-formula with time-varying covariates to address censoring, non-adherence, and competing events when estimating per-protocol effects.

A Latent Variable Approach for Causal Effect Estimation under Misclassified Treatment Assignment

Penn State University

Supervised by Dr. Lan Kong

03/2024 - 12/2024

- Proposed a latent variable approach that treats true treatment assignment as a latent variable for causal effect estimation, accounting for potential misclassification of treatment assignments.
- Decomposed the complete likelihood function into three components: the propensity score model, measurement error model, and outcome model, with parameters estimated using the expectation-maximization (EM) algorithm using R.
- Incorporated validation data and machine learning approach to enhance the measurement error modeling and 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, through simulation studies.

Robust Propensity Score Estimation via Loss Function Calibration

Penn State University

Supervised by Dr. Lan Kong

03/2023 - 03/2024

- Proposed robust propensity score estimation method under model misspecification by incorporating covariate imbalance into loss function of machine/deep learning methods, including neural networks and LASSO
- Conducted simulation studies with various model specifications to compare causal effect estimation using different propensity score methods and causal estimators (e.g., Horvitz-Thompson(HT), Hájek, doubly robust) using R and Python.
- Validated the robustness of the proposed method against both correctly specified and misspecified propensity score models, demonstrating a significant reduction in bias and RMSE.

High-dimensional Propensity Score Estimation via Outcome-Assisted Variable Selection for Real World Data (RWD)

Penn State University

Supervised by Dr. Lan Kong

06/2023 - 09/2024

- Extended the *loss function calibration* method to a high-dimensional setting by incorporating outcome-assisted variable selection for propensity score model.
- Extracted cohorts with high-dimensional baseline covariates to emulate clinical trial data using real-world data from the MarketScan Claims Database using SAS and SQL.
- Conducted plasmode simulations with the extracted real-world data to evaluate the proposed high-dimensional method.
- Demonstrated that the proposed method outperforms others (outcome adaptive LASSO, hdCBPS) in providing unbiased causal effect estimation.

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

Penn State University

Supervised by Dr. Shouhao Zhou

08/2022 - 08/2023

- Proposed a novel non-parametric approach to enhance estimation and hypothesis testing for transient survival data by conceptualizing state transitions as pseudo-competing events and reframing the analysis as a competing events problem.
- Calibrated the cumulative incidence function 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 a Shiny app and associated software paper for its application.

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

Columbia University

Supervised by Dr. Xiaoyu Che

06/2020-10/2020

- Constructed logistic regression and Cox hazard model to estimate the effect size for each biomarker; adjusted for multiple comparisons using the Hochberg step-up method.
- Implemented Adaptive LASSO, Random Forest, and XGBoosting algorithms as feature selection methods with Bootstrap for a robust predictive model.