

Yimeng Shang

☎ +1 (646) 704-5390

✉ yqs5519@psu.edu

🌐 <https://ys3298.github.io/>

EDUCATION

Pennsylvania State University, Hershey, PA
Ph.D. in Biostatistics (GPA: 4.0/ 4.0)

August, 2021 - Present

Columbia University, Mailman School of Public Health, New York, NY
Master of Science in Biostatistics (GPA: 4.0/ 4.0)

April, 2021

East China Normal University, Shanghai, China
Bachelor of Science in Mathematics and Applied Mathematics

June, 2019

University of California, Berkeley, Berkeley, CA
Berkeley International Study Program

August, 2017 – May, 2018

INTERNSHIP EXPERIENCE

Cytel, Boston, MA,

June, 2022 – August, 2022

Strategic Consulting Intern

- Proposed predictive variable/biomarker selection algorithm via knockoff filters
- Developed Shiny App to implement the proposed algorithm
- Supported clinical trial design simulation project and wrote the statistical analysis plan for FDA submission

Eli Lilly, Shanghai, China

September, 2018 – June, 2019

Data Science & Solution Intern

- Supported data management work in clinical trials including data cleaning and missing data query under supervision of China DSS team
- Participated in the discussion of the application of statistical methods in clinical trials
- Constructed quantitative analysis of possible interference risks during clinical trials and organization operation, developed an automatic web page with RShiny for reproducible monthly analysis to improve efficiency

RESEARCH EXPERIENCE

Robust Propensity Score Estimation with Application to MarketScan Real-World Data

Research assistant, supervised by Dr. Lan Kong, Penn State University

Fall, 2022 - Now

- Proposed robust propensity score (PS) estimation method by incorporating covariate imbalance into loss function of multiple machine learning methods
- Extend the method to high-dimensional claims data setting with outcome assisted variable selection method
- Conduct empirical simulation and plasmode simulation to compare ATE estimation using different PS methods
- Applied the proposed propensity score methods on gastrointestinal complications comparison using MarketScan real-world data and compare with clinical trials findings
- Manuscript in preparation

Graphical Representation and Statistical Test for Irreversible Time-dependent Ordinal Covariate for Survival Analysis and Competing Risk Analysis

Fall, 2021

Research assistant, supervised by Dr. Shouhao Zhou, Penn State University

- Developed graphical representation and statistical test for irreversible time-dependent ordinal covariate for survival and competing risk analysis by estimating and calibrating cumulative incidence function with inverse probabilistic weighting of stage transition
- Developed ShinyApp for application
- Manuscript in preparation

Variable Selection in Multiple Regression with Misclassification in the Response

Variable

Summer, 2020

Research assistant, supervised by Dr. Hua Shen, University of Calgary

- Proposed a new method combining EM algorithm and Adaptive LASSO for variable selection, parameter estimation, and prediction. Extended the method to be feasible with less assumptions by constructing two separate penalty term structure in EM algorithm and investigate misspecification problem
- Constructed multiple simulation studies under different settings, including considering different model assumptions, different dimensionalities, different regularization methods and different tuning parameter selection criteria to evaluate the robustness of new method
- Achieved consistent improvements in parameter estimation, variable selection and prediction compared to the Naïve method and Ad Hoc method under different settings

Statistical Analysis of High Dimensional Metabolomics Data in ASD

Summer, 2020

Research assistant, supervised by Dr. Xiaoyu Che, Columbia University

- Constructed both logistic regression and Cox hazard model to estimate the effect size; adjusted for multiple comparison using Hochberg step up method; Did 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 the purpose of building a robust predictive model.

PUBLICATIONS

Che, X., Roy, A., Bresnahan, M., **Shang, Yimeng** et al. Metabolomic analysis of maternal mid-gestation plasma and cord blood in autism spectrum disorders. *Mol Psychiatry* (2023). <https://doi.org/10.1038/s41380-023-02051-w>

Endres K M, 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]. *Journal of Emergency Nursing*, 2022.

Shechter, Ari & Chiuzan, Cody & **Shang, Yimeng** & Ko, Gavin & Diaz, Franchesca & Venner, Hadiyah & Shaw, Kaitlin & Cannone, Diane & McMurry, Cara & Sullivan, Alexandra & Rivera, Reynaldo & Vose, Courtney & Shapiro, Peter & Abdalla, Marwah. (2021). 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. *International Journal of Environmental Research and Public Health*. 19. 262. 10.3390/ijerph19010262.

Abdalla, Marwah & Chiuzan, Cody & **Shang, Yimeng** & Ko, Gavin & Diaz, Franchesca & Shaw, Kaitlin & McMurry, Cara & Cannone, Diane & Sullivan, Alexandra & Lee, Sung A & Venner, Hadiyah & Shechter, Ari. (2021). Factors Associated with Insomnia Symptoms in a Longitudinal Study among New York City Healthcare Workers during the COVID-19 Pandemic. *International journal of environmental research and public health*. 18. 10.3390/ijerph18178970.

SKILLS

Computer Skills: R, SAS, Python, MATLAB

Language Skills: Chinese Mandarin (Native), English (Fluent)