Yushu Shi, PhD

Department of Biostatistics
University of Texas MD Anderson Cancer Center
1400 Pressler Street
Houston, TX 77030

E-mail: yshi7@mdanderson.org
Phone: 202-367-6363
Office: FCT4.6009

Education and Training

2017 - Postdoc, University of Texas MD Anderson Cancer Center
Research focus: microbiome data, unsupervised clustering, feature selection
Advisor: Robert Jenq, Kim-Anh Do, Christine Peterson

2012 - 2017 Ph.D., Biotatistics, Medical College of Wisconsin
Dissertation: Weibull Mixture Models for Regression in the Context of Time-to-event Data
Advisor: Purushottam (Prakash) Laud

2008 - 2012 B.S., Statistics, Nankai University, Tianjin, China

Research Interests

Bayesian statistics, variable selection, microbiome data, survival analysis, unsupervised clustering

Experience

2017 - Postdoc at University of Texas MD Anderson Cancer Center, Department of Biostatistics

- Developed a Bayesian mixture model for unsupervised clustering
- Gained experience with Bayesian methods for feature selection
- Compared commonly used Beta diversity metrics for microbiome data
- Developed tool for microbiome data visualization
- Collaborated on clinical studies related to microbiome

2012 - 2017 Medical College of Wisconsin

Methodology Research

- Used Dirichlet process and dependent Dirichlet process mixture to model survival data with and without competing risks.
- Elicited low informative prior for Bayesian nonparametric mixture models

Teaching Assistant

- Graded homework.
- Answered students' questions.

CIBMTR(Center for international blood and bone marrow transplant research)

- Collaborated with clinicians on bone marrow transplant research

PCOR(Center for Patient Care and Outcomes Research)

- Collaborated with clinicians on breast cancer research with survival outcomes.

Jun-Aug 2016 Mayo Clinic, MN

Summer Intern

- Clustered longitudinal data for ARCAD colorectal cancer project.

Publications

1. **Y. Shi**, L. Zhang, R. Jenq, K.A. Do, and C. Peterson. Bayesian approaches for flexible and informative clustering of microbiome data. Submitted
2. **Y. Shi**, P. Laud and J. Neuner. A Dependent Dirichlet Process Model for Survival Data With Competing Risks. revised for *Lifetime Data Analysis*.
3. **Y. Shi**, M. Martens, A. Banerjee and P. Laud.(2019) Low information omnibus (LIO) priors for Dirichlet process mixture models. *Bayesian Analysis* **14**(3): 677–702.
4. J. Galloway-Peña, **Y. Shi**, C. Peterson, P. Sahasrabhojane, V. Gopalakrishnan, C. Brumlow, N. Daver, M. Alfayez, P. Boddu, M.A.W. Khan, J. Wargo, K.A. Do, R.Jenq, D. Kontoyiannis, and S. Shelburne.(2019) Gut microbiome signatures are predictive of infectious risk following induction therapy for acute myeloid leukemia. Accepted by *Clinical Infectious Diseases* .
5. S. Kamaraju, **Y. Shi**, E. Smith, A. Nattinger, P. Laud and J. Neuner.(2019) Are aromatase inhibitors associated with higher myocardial infarction risk in breast cancer patients? A Medicare population-based study. *Clinical Cardiology*. **42**: 93–100.
6. Z.D. Jiang, R. Jenq, N. Ajami, J. Petrosino, A. Alexander, S. Ke, T. Iqbal, A. DuPont, K. Muldrew, **Y. Shi**, C. Peterson, K.A. Do, and H. DuPont.(2018) Safety and preliminary efficacy of orally administered lyophilized fecal microbiota product compared with frozen product given by enema for recurrent Clostridium difficile infection: A randomized clinical trial. *PLOS ONE*. 13(11):1-12, 11
7. J. Neuner, **Y. Shi**, A. Kong, S. Kamaraju, E. Smith, A. Smallwood, P. Laud and J. Charlson.(2018) Fractures in a nationwide population-based cohort of users of breast cancer hormonal therapy. *Journal of Cancer Survivorship*. 12: 268.
8. Y.C. Cheng,**Y. Shi**, M.J. Zhang, R. Brazauskas, M. Hemmer, M. Bishop, Y. Nieto, E. Stadtmauer, L. Ayash, R. Gale, H. Lazarus, L. Holmberg, M. Lill, R. Olsson, B.M. Wirk, M. Arora, P. Hari and N. Ueno.(2017) Long-Term Outcome of Inflammatory Breast Cancer Compared to Non-Inflammatory Breast Cancer in the Setting of High-Dose Chemotherapy with Autologous Hematopoietic Cell Transplantation. *Journal of Cancer*. 8(6):1009-1017.
9. A. Biggers, **Y. Shi**, J. Charlson, E. Smith, A. Smallwood, A. Nattinger, P. Laud, and J. Neuner.(2016) Medicare D Subsidies and Racial Disparities in Persistence and Adherence With Hormonal Therapy. *Journal of Clinical Oncology*. 34(36):4398-4404.

Presentations and Posters

A Bayesian Approach for Flexible Clustering of Microbiome Data, poster, Mar 24th, 2019, ENAR, Philadelphia, PA

A Low Informative Prior for Dirichlet Process mixture of Weibull Distributions, poster, Mar 6th, 2016, ENAR, Austin, TX

Performance Determinants of Unsupervised Clustering Methods for Microbiome Data, presentation, Apr 10th, 2019 microbiome working group meeting

Software

R package `DPWeibull`: Dirichlet process Weibull mixture model and dependent Dirichlet process Weibull mixture model for survival data with and without competing risks.
Available at https://cran.r-project.org/web/packages/DPWeibull/index.html
R package `MicrobiomeCluster`: Basic functions for microbiome data manipulation.
Available at https://github.com/YushuShi/MicrobiomeCluster.git
R package/ Matlab code `BayesianMicrobiome`: Select important OTUs or phylogenetic tree nodes and cluster observations.
Available at https://github.com/YushuShi/BayesianMicrobiome.git
R package `AdjustedPCoA`: Covariate adjusted PCoA plot.
Available at https://github.com/YushuShi/BayesianMicrobiome.git
R shiny app in https://chinawokee.shinyapps.io/covariateadjpcoa/

Professional Training

Workshop
Big Data, Data Science and Deep Learning for Statistician, 2019 ENAR
Certificate
Advanced programming for SAS 9
SAS Clinical Trials Programming Using SAS 9

Professional Service and Membership

Journal Review
Biometrics (joint review with Christine Peterson)
Membership
- American Statistical Association (ASA)
- Eastern North American Region International Biometric Society (ENAR)

Computer Skills

SAS, R, Rcpp, MATLAB, basic linux