Research Statement

Kenny Jing Xu

**Introduction**

I am a well-rounded statistician with broad interest and expertise in both methodology and applications. My research interest primarily focuses on proposing novel statistical methodologies either in experimental design or data analysis that can be applied to scientific research. I believe in wide dissemination of research, in conjunction with open-source software as and when appropriate. My research goal is not only to publish statistical methodology papers in high rank journals, but also to publish the corresponding R packages in the Comprehensive R Archive Network (CRAN).

**Previous Research**

My PhD (Mentor: A/Prof Jun Ma) research at Macquarie University, Australia, was about fitting Cox proportional hazard models under dependent censoring using maximum penalized likelihood (MPL) (Xu, et al (2018)). Dependent censoring was captured by a copula function. The non-parametric baseline hazard function was approximated by a spline function, which was estimated by a positive constrained algorithm called “Multiplicative Iterative (MI) algorithm”. The consistency and asymptotic normality properties of the MPL estimators were proved. Simulation studies were conducted to compare it with the method of Chen (2010). The proposed method had been applied to a dementia study called “Prospective Research in Memory Clinics” (PRIME). It was utilized to select the predictors for dementia patients to be institutionalized (Brodaty, et al (2014)). The corresponding R package “survivalMPLdc” can be downloaded from Github.

As a post-doctoral research fellow (Mentor: A/Prof Bibhas Chakraborty) at Duke-National University of Singapore Medical School (Duke-NUS), Singapore, I started research on developing novel statistical methodologies in the cutting-edge arena of personalized medicine, i.e. dynamic treatment regimens (DTRs) for chronic diseases. We proposed a possible set of scientifically meaningful DRTs for periodontitis, and also proposed a two-stage sequential multiple assignment randomized trial (SMART) design called “SMARTp” to study them. We further developed a statistical analysis plan and a novel cluster-level sample size calculation method that factors in typical features of periodontal responses, i.e. non-Gaussianity, spatial clustering, and non-random missingness of teeth. The expected values and variances of the sample means of a DTR are derived by the inverse probability weighting method and the method of moments. The consistency and asymptotic properties of the proposed estimator were proved. Simulation studies were conducted to investigate the finite-sample performance of the proposed sample size formulas. The resulting manuscript (Xu, et al (2019)) has been published in Biometrical Journal. The corresponding R package “SMARTp” is available in CRAN while the corresponding R shiny application “SMARTp-SS” is shared as web-link (<https://kennyxu.shinyapps.io/SMARTp-SS/>).

**Current Research**

During my post-doctoral work at Duke-NUS, I also worked on a second thread of research on mobile health (mHealth), particularly in the context of behavioural sciences. Specifically, this work develops a novel version of micro-randomized trial (MRT) design (Klasnja, et al (2015) and Liao, et al (2016)) with the corresponding sample size calculator. We call the proposed design “Multi-Level Micro-Randomized Trial” (MLMRT). This design will be applied to a real mHealth study called “DIAMANTE”, to be held at the San Francisco Health network, California, USA. DIAMANTE is a free mobile application used in the Diabetes and Mental Health Adaptive Notification Tracking and Evaluation (DIAMANTE) study to understand how contextually tailored text messaging programs can be used to increase physical activity among people with co-morbid diabetes and depression. The purpose of this study is to learn the best ways to develop and deliver text messages to encourage individuals to do regular physical activity in daily life. Simulation studies show that given the calculated sample sizes, the Monte Carlo estimated powers and coverage probabilities are very close to the corresponding nominal values. We have completed a manuscript for the proposed method and submitted it to arXiv (Xu et al (2020)). The corresponding R shiny Application of the sample size calculator is shared as a website (<https://kennyxu.shinyapps.io/mlmrt_shinyapps/>).

**Future Research Plan**

**mHealth**

I will be working closely with A/Prof Chakraborty to submit our MLMRT manuscript. We will also work closely with our American collaborators, to provide statistical support to the DIAMANTE study in order to bring it to completion. Our role will contribute to not only the experimental design and sample size calculation, but also data analysis.

We aim to develop novel statistical methods based on analysing DIAMANTE dataset, complete a manuscript based on the proposed analysis method and submit it to a high

rank statistical journal. We also aim to complete a manuscript for the R package “MLMRT” and submit it to Journal of Statistical Software or R Journal.

The proposed MLMRT design was not only applied to DIAMANTE, but also applied to an another study called “Stay Well at Home: a Text-messaging Study Social Distancing”. The purpose of this study is to develop and examine the supportive text-messages that used to enhance positive mood for people cope with the stress and anxiety of COVID-19 social distancing. The study cohort targets low-income, vulnerable populations across the United States. The study protocol is under preparation. In the similar plan for the DIAMANTE study, we aim to develop novel statistical method through data analysis.

**DTR**

I will continue to collaborate with A/Prof Chakraborty, Prof Dipankar Bandyopadhyay and our oral health collaborators for both methodology and application research of DTR on periodontitis. For example, the method of Xu, et al (2020) can be extended to consider covariates using Q-learning approach. Similar to Ertefaie, et al (2015), we can also consider selecting the optimal DTR as the primary criterion to drive the sample size calculation for our “SMARTp” design.

**Survival Analysis**

I will collaborate with A/Prof Ma for the research on survival analysis. We aim to submit the R package for Xu et al (2018) to CRAN. I am also interested in extending this method to other types of semi-parametric hazard models, e.g., accelerated failure time model, additive hazard model, or time dependent covariates hazard models, using the PRIME dataset. Similar to Emura and Chen (2016), our method can be extended for gene selection for cancer survival data under dependent censoring.

**Reference**

**Xu, J**., Yan, X., Figueroa, C., Williams, J.J. and Chakraborty, B. (2020). Multi-Level Micro-

Randomized Trial for Detecting the Proximal Effect of a Mobile Application Messages on Physical Activity in Diabetes and Depression. arXiv, stat.ME, 2007.13741.

**Xu, J**., Bandyopadhyay, D., Mirzaeic, S., Michalowicz, B. and Chakraborty, B. (2020). *SMARTp: A SMART design for non-surgical treatments of chronic periodontitis with spatially-referenced and non-randomly missing skewed outcome.* Biometrical Journal, https://doi.org/10.1002/bimj.201900027 .

Klasnja, P., Smith, S., Seewald,N.J., Lee, A., Hall, K., Luers, B., Hekler, E.B., Murphy, S.A. (2019) Efficacy of Contextually Tailored Suggestions for Physical Activity: A Micro-randomized Optimization Trial of HeartSteps, Annals of Behavioral Medicine. 53:6, 573–582.

**Xu, J**., Ma, J., Connors, M.H. and Brodaty, H. (2018) Proportional hazard model estimation under dependent censoring using copulas and penalized likelihood. Statistics in Medicine, 37:14, 2238-2251.

Emura, T. and Chen, Y.H. (2016). Gene selection for survival data under dependent censoring: A copula-based approach. Statistical Methods in Medical Research, 25(6), 2840–2857.

Liao P, Klasnja P, Tewari A, Murphy SA. (2016) Sample size calculations for micro-randomized trials in mHealth. Statistics in Medicine. 35:12, 1944–1971.

Klasnja P, Hekler E, Shiffman S, Boruvka A, Almirall D, Tewari A, Murphy S. Micro-Randomized Trials: An Experimental Design for Developing Just-in-Time Adaptive Interventions. *Health Psychology: Official Journal of the Division of Health Psychology, American Psychological Association* 2015; 34:1220–1228.

Ertefaie, A., Wu, T., Lynch, K., and Nahum-Shani, I. (2015). Identifying a set that contains the best dynamic treatment regimes. *Biostatistics* 17, 135–148.

Brodaty, H., Connors, M.H., **Xu, J**., Woodwards, M., Ames, D. (2014) Predictors of Institutionalization in Dementia: A three Year Longitudinal Study. Journal of Alzheimer’s Disease, Vol. 40, 1, 221-226.

Chen, Y.H. (2010). Semiparametric marginal regression analysis for dependent competing risks under an assumed copula. Journal of the Royal Statistical Society, Series B, 72, 235-251.