Workshop on Statistical Methods for Children's Health Research
September 18, 2015, 1:30-5:00PM
1690 SPH I (Lane Auditorium)

1415 Washington Heights
School of Public Health

Program

Sessions and Schedule (1:30-5:00PM)	
Session I: Statistical Models for Complex Longitudinal Trajectories Chair: Mathieu Bray, Department of Biostatistics, University of Michigan	
1:30-2:10 + 5 min Q&A	**Bin Zhu** (National Cancer Institute, NIH) Bayesian functional data modeling for heterogeneous volatility
2:15-2:55 + 5 min Q&A	**Shujie Ma** (Department of Statistics, University of California, Riverside) Evaluating alterations in longitudinal regression coefficients directed by toxicant mixtures
3:00-3:15	Coffee Break
Session II: Statistical Methods for Precision Medicine Chair: Peter Song, Department of Biostatistics, University of Michigan	
3:15-4:05 + 10 min Q&A	**Heping Zhang** (Department of Biostatistics, Yale University) Decision trees for precision medicine
Session III: Statistical Methods for Variable Selection Chair: Emily Hector, Department of Biostatistics, University of Michigan	
4:15-4:35	**Wen Wang** (Department of Biostatistics, University of Michigan) Quasi-likelihood under the dependent model criterion
4:35-4:55	**Lu Tang** (Department of Biostatistics, University of Michigan) Fused lasso approach in regression coefficients clustering
4:55-5:00	Q&A

Presentation Abstracts

Session I: Statistical Models for Complex Longitudinal Trajectories

Bin Zhu (1:30-2:15PM)

Bayesian functional data modeling for heterogeneous volatility

Many methods have been proposed to model the trend of functional data, but less emphasis is put on characterizing variability among volatilities of individual functions. In particular, certain individuals exhibit erratic swings in their trajectory while other individuals have more stable trajectories. There is evidence of such volatility heterogeneity in blood pressure trajectories during pregnancy, for example, and reason to suspect that volatility is a biologically important feature. To capture volatility heterogeneity, we proposed a stochastic volatility regression model, allowing volatilities dependence upon covariates. We applied the proposed method and identified obesity and preeclampsia being significantly associated with higher volatility of blood pressure trajectories during pregnancy in the Healthy Pregnancy, Healthy Baby study.

Shujie Ma (2:15-3:00PM)

Evaluating alterations in longitudinal regression coefficients directed by toxicant mixtures

We propose a new linear mixed effects model for the analysis of longitudinal growth data. This new class of longitudinal models enables us to evaluate adverse effects of exposure to toxicant mixtures on somatic growth trajectories. The proposed statistical model is motivated and illustrated by an analysis of child growth data that aims to assess alterations in growth rates incurred by mother's exposure to endocrine disrupting compounds during pregnancy. In particular, using the proposed model and algorithms we can assess if and how the rate of growth is modified by exposures to mixtures of toxicants (e.g. PBA and phthalates). Difficulty of statistical analysis arises from the fact that exposure to multiple toxicants is simultaneous in reality and that most of toxic agents are of small size in their effects and thus the traditional statistical methods fail to detect statistical significance. In this project we introduce a new longitudinal data modeling strategy that invokes the idea of principal component analysis in the formation of regression coefficients, termed as index coefficients, in that low-effect toxicants are combined into possibly strong-effect toxicant groups. Statistical estimation and inference in this new model is challenging because it contains nonlinear interactions between the toxicant groups and covariates of interest (e.g. age or time). This talk is based on the joint work with Peter Song from Department of Biostatistics, University of Michigan.

Session II: Statistical Methods for Precision Medicine

Heping Zhang (3:15-4:15PM)

Decision trees for precision medicine

Double-blind, randomized clinical trials are the preferred approach to demonstrating the effectiveness of one treatment against another. The comparison is, however, made on the average group effects. While patients and clinicians have always struggled to understand why patients respond differently to the same treatment, and while much hope has been held for the nascent field of predictive biomarkers (e.g. genetic markers), there is still much utility in exploring whether it is possible to estimate treatment efficacy based on demographic and baseline variables including biomarkers. To address this issue, we focused on a concept of the relative effectiveness of treatments that is of particular importance in precision medicine. The method can identify groups of patients that are more likely to respond one treatment than the other, in contrast to the tradition approach that searches for a superior treatment in a larger population. We developed an automated algorithm to construct decision trees and performed extensive simulation to evaluate our algorithm. We analyzed data from clinical trials to illustrate the practical potential of our method.

Session III: Statistical Methods for Variable Selection

Wen Wang (4:15-4:35PM)

Quasi-likelihood under the dependent model criterion

Generalized estimating equation (GEE) gained its popularity in application because its minimum model specification requirement. Quasi-likelihood under independent model criterion (QIC) was proposed to use marginal quasi-likelihood as model selection criterion in GEE. The purpose of this paper is to explore whether using quasi-likelihood or likelihood's counterpart in GEE rather than marginal quasi-likelihood as model selection criterion is possible. As a result, quasi-likelihood under the dependent model (QDC) was introduced. Simulation study showed advantage of QDC compared to QIC in covariate selection.

Lu Tang (4:35-4:55PM)

Fused lasso approach in regression coefficients clustering

The trend of making research data public with article publications has driven statistical studies toward a more collaborative and collective environment. As data sets of related studies become more easily accessible, combining data sets collected from similar studies over different subjects are routinely undertaken in practice to achieve a larger sample size and higher power. A major challenge arising from data integration pertains to data heterogeneity in terms of inter-study population, study coordination, or experimental protocols. Ignoring such data heterogeneity in data analysis may result in biased estimation and misleading inference. Traditional techniques of remedy to data heterogeneity include the utility of interactions and random effects, which are inferior to achieving desirable statistical power or providing an intuitive interpretation, especially when the data integration involves a large number of smaller data sets. In this paper, we propose a regularized fusion method that allows us identify and merge inter-study homogenous parameter clusters in regression analysis, without the use of hypothesis testing approach. Using the fused lasso in parameter fusion, we establish a computationally efficient procedure to deal with large-scale integrated data in that the existing statistical software is readily applicable. Incorporating the estimated parameter ordering in the fused lasso facilitates computing speed with no loss of statistical power.