A. (CLINICAL) Variable Selection in Semiparametric Survival Models for Correlated Failure-Time Data with Long-Term Survivors

Speakers: Abdullah Al Masud[1], Zhangsheng Yu[2], and Wanzhu Tu[3]

1. AbbVie Inc., 1 North Waukegan Road, North Chicago, IL 60064
2. Department of Bioinformatics & Biostatistics, Shanghai Jiaotong University, China
3. Department of Biostatistics, Indiana University School of Medicine, Indianapolis, IN 46202

Statistical models for analyzing correlated failure-time data with long-term survivors have two components: A logistic model that depicts the rate of long-term survivors, and a survival model that describes the distribution of the survival outcomes. Model selection methods are needed for screening the important variables and for determining the functional forms of the potentially nonlinear effects. In this research, we propose a regularized regression method for variable selection, in both modeling components. The method accommodates repeated events as well as long-term survival. The proposed procedure is based on the Least Absolute Shrinkage and Selection Operators (LASSO) and it can be used to assist the determination of the composition of complex models. For computation, we decompose each variable into linear and nonlinear parts, where the nonlinear part is estimated by coefficients of cubic-B spline bases. The procedure differs from the existing variable selection methods in its ability to discover hidden nonlinear structures in the independent variables. To implement, we use an Expectation Maximization (EM) algorithm. We conduct extensive simulation studies to evaluate the operating characteristics of the proposed methods. We illustrate the method by analyzing data from an observational study of sexually transmitted infections (STIs).