

# Principled Sure Independence Screening for Cox Models with Ultra-high-dimensional Covariates

*Yi Li*

*Harvard University*

yili@hsph.harvard.edu

## Abstract

It is rather challenging for current variable selectors to handle situations where the number of covariates under consideration is ultra-high. Consider a motivating study of clinical trials of bortezomib for the treatment of multiple myeloma, where overall survival and expression levels of 44760 probesets, encompassing more than 22000 genes, were measured for each of 188 patients with the goal of identifying genes that predict survival after treatment. This dataset defies analysis even with regularized regression. Some remedies have been proposed for the linear model and for generalized linear models, but there are few solutions in the survival setting and, to our knowledge, no theoretical support. Furthermore, existing strategies often involve tuning parameters that are difficult to interpret. In this paper we propose and theoretically justify a principled method for reducing dimensionality in the analysis of censored data by selecting only the important covariates. Our procedure involves a tuning parameter has a simple interpretation as the desired false positive rate. We apply the proposed procedure to analyze the aforementioned myeloma study and identify biologically important and predictive genes.