

## Use of resampling methods for the evaluation of high-dimensional survival risk models

Richard Simon\*  
rsimon@mail.nih.gov

Jyothi Subramanian\*  
jyothi.subramanian@nih.gov

---

Developments in whole genome biotechnology have stimulated statistical focus on prediction methods. We review here methodology for classifying patients into survival risk groups and for using cross-validation to evaluate such classifications. Measures of discrimination for survival risk models include separation of survival curves, time-dependent ROC curves, and Harrell's concordance index. For high dimensional data applications, however, computing these measures as re-substitution statistics on the same data used for model development results in highly biased estimates. Most developments in methodology for survival risk modeling with high dimensional data have utilized separate test dataset for model evaluation. In many applications, however, the data available is too limited for effective division into training and test sets. This talk will describe how to compute cross-validated estimates of survival distributions for predicted risk groups and how to compute cross-validated time-dependent ROC curves. We also discuss evaluation of the statistical significance of a survival risk model and evaluation of whether high dimensional genomic data adds predictive accuracy to a model based on standard covariates. We will also describe our work comparing different resampling techniques for their ability to estimate the accuracy of risk prediction models. Our study showed that accuracy estimates for popular methods such as sample splitting and leave-one-out cross-validation have a higher mean square error than for other methods.  $k$ -fold cross-validation with  $k = 5$  or  $10$  was seen to provide a good balance between bias and variability for a wide range of data settings and should be more widely adopted in practice.

---

\*Biometric Research Branch, National Cancer Institute, 6130 Executive Plaza, Room 8134, Rockville, MD 20852, USA.