Q 11. CHOOSING WHAT VARIABLES MATTER IN RISK STRATIFICATION FOR HEPATECTOMIES USING A MACHINE LEARNING TECHNIQUE

Presenter: Maria Urdaneta Perez MD | Oklahoma University College of Medicine
MG Urdaneta, T Garwe, K Stewart, Z Sarwar, KT Morris

Background: Classification and regression trees (CART) are a group of machine learning techniques used to identify variables with the best discrimination for a specific outcome, optimizing accuracy in risk modeling. Here we used the CART method recursive partitioning in screening numerous potential predictors of mortality or failure to discharge home after hepatectomy. The goal was to identify a group of key predictors through risk stratification and to determine an optimal threshold for the dichotomization of these variables to maximize accuracy and simplicity in subsequent predictive modeling.

Methods: Data were obtained from the NSQIP PUF 2005-2017 covering hepatectomies. Two classification tree models were developed, one estimating probability of 30-day mortality, the other failure to discharge home. Twenty-eight covariates were considered including: demographics, comorbidities, pre-operative labs, procedure type and functional status. Although indication for resection was also a considered variable, due to a very broad range of diagnoses obtained from the NSQIP dataset, all indications for hepatectomies were included. Continuous predictor variables were converted to dichotomous variables after determining inflection points with the best discriminating ability for the two models. These were then set as our cut-offs. Multiple imputation was used to complete missing lab values. Model discrimination was assessed using the area under the receiver operating characteristic curve (AUC).

Results: Crude 30-day mortality was 1.8% (654/37,141). Among survivors, 1,808 (6.2%) of 29,106 patients were not discharged home (discharge location available 2011-2017). CART results estimated probabilities of death ranging from 0.4% to 27%. The estimated probability of not being discharged home ranged from 2-49%. Starting from 28, CART identified 6 variables for predicting mortality (AUC=0.76) and 7 variables for predicting failure to discharge home (AUC=0.74) with minimal improvement in discrimination with the addition of more variables. For the mortality model the following variables and cut-off points were identified: procedure type dichotomized as right/trisegmentectomy vs partial/left, age < or > 70, albumin < or > 2.5g/dL, bilirubin < or > 3g/dL, platelets < or > 100K and presence or absence of diabetes. For the discharge not home model, the identified variables and cut-off points were: Age < or > 70, functional status dichotomized as partial/total dependence vs independent, albumin < or > 2.5g/dL, procedure type dichotomized as right/trisegmentectomy vs partial/left, presence or

absence of hypertension, hematocrit < or > 30% and presence of COPD. Future efforts may be conducted to consider diagnosis-specific models by using NSQIP hepatectomy targeted data.

Conclusion: Recursive partitioning identified a minimally sufficient number of covariates most strongly associated with two important hepatic resection outcomes which would have been difficult to accomplish with traditional logistic regression models. This screening technique can guide subsequent predictive modeling and potentially result in a simpler hepatic resection outcome risk score development.