SO-22 **Importance order of risk factors for mortality in colon cancer stage II: A multivariate adaptive regression spline (MARS) analysis**

J. Fabregas

Harvard School of Public Health, Boston, United States

Background: 52,980 deaths from colon cancer are expected in the US in 2021. Early stages tend to fare better, but recurrence and death are possible. The use of adjuvant chemotherapy in Stage II colon cancer is considered on an individual basis. To make the decision, clinicians evaluate the presence of high-risk factors. Their relative order of importance for mortality risk is unknown.

Methods: The National Cancer Database was queried for patients with colon cancer stage II. It includes approximately 70% of cancer cases in the US. 70,074 patients were included in the final analysis. MARS, a machine learning technique, was used to construct a model to predict survival in patients with Stage II colon cancer and to determine variable importance. Data were partitioned into training, validation and testing groups at a 3:1:1 ratio. Outcome variable was death. Based on previous literature, we included the following variables as predictors to train the model: tumor grade, T staging, number of lymph nodes examined, carcinoembryonic antigen (CEA) level, perineural and lympho-vascular invasion. Microsatellite instability and bowel perforation data were not available. To determine the importance of each variable, we used a relative importance score (0-100%) generated in the statistical package. An area under receiving operating characteristic (ROC) curve in the training and testing dataset was generated to determine fit of the model. A generalized cross-validation score (GCV) was used to select optimal number of base functions. Statistical analysis was conducted in Salford Predictive Modeler v. 8.3.

Results: Relative importance scores of the variables were as follows: 1. Number of lymph nodes examined (100%), 2. CEA level (82.2%), 3. T4 staging (50.34%), 4. Tumor grade (31.31%). 5. Lympho-vascular invasion (25.44%). Perineural invasion did not contribute to the model. The MARS model generated an ROC curve for the training dataset 0.6185 and for the testing dataset 0.6179. GCV score was 0.19204 and suggested 7 basis functions. The model generated two knots for number of lymph nodes examined, at 14 and 27 lymph nodes. Patients with less than 14 lymph nodes examined were most likely to die. This relationship persisted but tended to plateau between 14 and 27 lymph nodes examined. For CEA level, one knot was defined at 38. The value of CEA elevation was linearly correlated with death until 38 ng/ml, after this number the maximum increase in the risk of death was almost reached.

Conclusions: According to our MARS model, the three most important risk factors associated with mortality in patients with colon cancer stage II are: 1. Less than 14 lymph nodes examined, 2. Elevated CEA level, and 3. T4 staging (vs T3), in that order. CEA level could be considered to be incorporated in national guidelines as an important risk factor to guide the administration of adjuvant chemotherapy. A prospective validation of CEA level as a mortality marker in patients with stage II colon cancer is needed.