Computational Risk Model for Predicting 2-Year Malignancy of Pulmonary Nodules using Demographic and Radiographic Characteristics

Objective: Demographic and radiographic features may be able to characterize whether a pulmonary nodule is malignant. The purpose of this study was to develop a risk model that can predict malignancy of pulmonary nodules. We hypothesized that both conventional and machine learning-based models would be superior to provider opinion in determining the risk of malignancy.

Methods: All patients enrolled in the prospective NAVIGATE trial (NCT02410837) in the United States were included, and patients were excluded for insufficient follow-up or missing study variables. The outcome of malignancy was determined after 2-years of follow-up. Multivariate analysis of demographic data and radiographic characteristics associated with malignancy was performed using stepwise-backwards selection (?<0.1). A random forest model was also constructed using the 'randomForest' package in R version 4.2.2. The accuracy of each model was determined using receiver operating characteristics (ROC) analysis and reported as an area under the curve (AUC); this was compared to the AUC of pre-biopsy provider opinion of malignancy.

Results: Among 984 patients in the analysis, 735 (74.7%) were diagnosed with malignancy. Factors associated with malignancy in multivariable regression were: age (OR 1.03, p<0.001), lesion size (OR 1.02, p=0.001), exposure to diesel fumes (OR 0.54, p=0.023), PET positivity (OR 2.60, p<0.001), history of pneumonia (OR 0.55, p=0.013), bronchus sign present (OR 1.41, p=0.042), personal history of cancer (OR 1.54, p=0.016), personal history of lung cancer (OR 2.11, p=0.020), and duration of tobacco use (OR 1.01, p<0.001). Random forest analysis was performed on the same group of patients. Variable importance in the model is shown in Figure 1, with the 3 most important variables, as determined by mean decrease in Gini Index, being subject age, largest lesion size, and total pack-years of tobacco use. The AUC for the multivariable model was 0.739, for the random forest model was 0.696, and for provider opinion was 0.828 (p<0.001).

Conclusions: Mathematical models of malignancy risk in pulmonary nodule are inferior to provider opinion. This study questions the ability of "bedside" prognostication tools using conventional radiographic and demographic characteristics to provide additional insight into patient care. Despite the importance of radiographic assessment, expert clinician evaluation of malignancy is paramount.

Kunaal Sarnaik (1), Allison Gasnick (1), Aria Bassiri (2), Craig Jarrett (2), Jillian Sinopoli (2), Leonidas Tapias Vargas (2), Philip Linden (2), Christopher Towe (2), (1) Case Western Reserve University School of Medicine, Cleveland, OH, (2) University Hospitals Cleveland Medical Center, Cleveland, OH