Aim: Lung cancer is the most common cancer in the world, more than 1,800,000 people died as a result of lung cancer in 2012. The most difficult challenge is to identify the symptoms in the early stage of lung cancer. Over half a million people could be saved each year if lung cancer can be detected in the early stage. The purpose of this study is to provide a method to predict and evaluate the early lung cancer risk.

Methods: A total of 345,600 people (2010-2013) including 9,500 lung cancer patients and 336,100 normal people were involved in the study. The data used in the study included demographic, CBC (Complete Blood Count), CMP (Complete Metabolic Panel), lipids and urinalysis data, a total of 48 data points. Both logistic analysis and discriminant analysis were used to identify the significant factors and to build the lung cancer risk prediction model and the significance level was set at p < 0.05. SAS was used as the primary statistical analysis tool. All the data were pulled out from the MS SQL database.

Results: The analysis results showed that there were 31 parameters that can significantly distinguish normal people from lung cancer patients and the accuracy of the prediction model was 95.5%. The top 10 parameters selected by the prediction model were gamma glutamyl transpeptidase (GGT), urobilinogen (URO), red cell distribution width (RDW), total bilirubin (TBIL), total cholesterol (CHO), percentage and absolute differential counts of monocytes, platelet (PLT), percentage of lymphocytes and red blood cell count (RBC). The prediction model was verified by a total of 120,008 people (2014) including 9,931 cancer patients and 110,077 normal people. The accuracy of the verification was 96.8%.

Conclusions: Routine blood and urine test results can be used to predict the probabilities of the early lung cancer risk and the accuracy of the prediction is over 95%. The results of this study could save millions of lives across the world.

Disclosure: All authors have declared no conflicts of interest.