

Computer-assisted Flagging of Never Smokers at High Risk of NSCLC in a Large US-based HMO using the LungFlag Model

Eran N. Choman¹, Alon Lanyado¹

¹ Medial EarlySign, Hod Hasharon, Israel.

Background

Smoking is considered to be the major cause of lung cancer, but lung cancer is not just a smokers’ disease. The prevalence of lung cancer in never-smokers is gradually rising with around 20% of lung cancers in the UK and US occurring in people who have never smoked^{1,2}. This figure rises to around 50% in some Asian countries^{3,4}.

Methods

A machine-learning algorithm based on routine EHR and laboratory data, previously developed and validated on an ever-smoker population⁵ was used to evaluate the accuracy in detection of Non-Small Cell Lung Cancer (NSCLC) among individuals who never smoked. The cohort was from a large US health system and included 509 case patients with NSCLC and 50,001 contemporaneous NSCLC-free controls. We compared the performance of two risk prediction models, LungFlag and the PLCOm2012 model adapted to EHR data (mPLCOm2012).

Results

Data were analyzed using the area under the receiver operating characteristic curve (AUC), positive predictive value (PPV), and diagnostic odds ratio (OR) as measures of model performance for the age group 40 and above. The risk predictors were calculated for multiple time windows prior to the diagnosis date (Dx) using cut-offs yielding specificities of 90%, 95%, 97% or 99%.

Conclusion

By using available information existing in the EHR, the model demonstrated high accuracy (OR>6) in early detection of NSCLC among never-smokers with data going back up to 24 months before diagnosis.

Furthermore, LungFlag creates an opportunity to carry out case finding in a population with growing rates of lung cancer that is currently not offered any screening, yet additional local validations are recommended.

Contact

Eran N Choman
eran@earlysign.com
+972 52 5611154
www.earlysign.com

Table 1: All NSCLC, Never Smokers, Ages 40+, 3-6 Months Before Dx

Model	AUC	Specificity Level							
		99%		97%		95%		90%	
		PPV %	OR	PPV %	OR	PPV %	OR	PPV %	OR
LungFlag	0.805* [0.773,0.840]	3.7* [2.3,5.6]	12.0 [7.1,19.5]	2.3 [1.7,3.3]	8.3 [5.7,12.7]	2.2* [1.7,2.9]	8.8* [6.5,13.5]	1.7* [1.4,2.10]	8.7* [6.5,12.3]
mPLCOm2012	0.760 [0.725,0.804]	2.2 [1.1,3.7]	7.6 [3.4,13.1]	1.8 [1.2,2.5]	7.0 [4.2,10.3]	1.3 [0.9,1.9]	5.2 [3.5,8.0]	1.1 [0.8,1.5]	5.1 [3.5,7.5]

* Statistically significant difference (P-value<0.05)

Table 2: All NSCLC, Never Smokers, Ages 40+, 6-9 Months Before Dx

Model	AUC	Specificity Level							
		99%		97%		95%		90%	
		PPV %	OR	PPV %	OR	PPV %	OR	PPV %	OR
LungFlag	0.821* [0.793,0.850]	2.6 [1.1,4.3]	8.2 [3.6,14.1]	2.3 [1.5,3.1]	8.0 [5.1,11.6]	2.0* [1.5,2.5]	7.7 [5.2,10.5]	1.4 [1.2,1.8]	6.4 [4.9,8.9]
mPLCOm2012	0.778 [0.740,0.814]	2.3 [1.2,3.9]	7.3 [3.3,12.7]	1.9 [1.3,2.6]	6.7 [4.2,10.0]	1.5 [1.0,1.9]	5.4 [3.4,7.6]	1.4 [1.0,1.6]	6.0 [4.1,8.0]

* Statistically significant difference (P-value<0.05)

Table 3: All NSCLC, Never Smokers, Ages 40+, 12-24 Months Before Dx

Model	AUC	Specificity Level							
		99%		97%		95%		90%	
		PPV %	OR	PPV %	OR	PPV %	OR	PPV %	OR
LungFlag	0.802* [0.783,0.820]	3.9 [3.1,5.1]	11.0 [8.3,15.8]	3.2* [2.7,3.7]	10.0* [8.0,12.2]	2.5* [2.2,3.0]	8.3* [7.0,10.3]	1.9* [1.7,2.1]	6.9* [5.9,8.6]
mPLCOm2012	0.756 [0.735,0.779]	3.4 [2.5,4.1]	9.5 [6.6,12.8]	2.6 [2.1,3.1]	7.6 [5.9,9.4]	1.9 [1.6,2.3]	5.7 [4.5,7.0]	1.5 [1.3,1.7]	4.9 [4.0,5.9]

* Statistically significant difference (P-value<0.05)

References

¹ Tindle HA, Stevenson Duncan M, Greevy RA, et al. 2018. Lifetime smoking history and risk of lung cancer: results from the framingham heart study. J Natl Cancer Inst 110(11): 1201-07

² Cufari ME, Proli C, De Sousa P, et al. 2017. Increasing frequency of non-smoking lung cancer: presentation of patients with early disease to a tertiary institution in the UK. Eur J Cancer 84: 55-59

³ Yang P. 2021. PS01.02 National Lung Cancer Screening Program in Taiwan: the TALENT Study. J Thorac Oncol 16(3): S58

⁴ Yung-Hung L, Chao-Hua C, MD, Chih-Hsi SK, et al. 2021. Lung Cancer in Republic of China. J Thorac Oncol 16(4): 519-527

⁵ Gould MK, Huang BZ, Tammemagi MC, et al. 2021. Machine Learning for Early Lung Cancer Identification Using Routine Clinical and Laboratory Data. AJRCCM 204(4).