FPN: **1561P**

TARGETED SCREENING METHODOLOGIES TO SELECT HIGH RISK INDIVIDUALS: LUNGFLAG™ PERFORMANCE IN ESTONIAN LUNG CANCER SCREENING PILOT

Tanel Laisaar[1,2], Eran Choman[3], Alon Lanyado[3], Eitan Israeli[3], Andre Koit[4], Kaja-Triin Laisaar[2]

[1]Lung Clinic, Tartu University Hospital, Tartu, Estonia, [2]Institute of Family Medicine and Public Health, University of Tartu, Tartu, Estonia; [3]Medial EarlySign, Hod Hasharon, Israel; [4]Roche Eesti OÜ, Tallinn, Estonia

Background

Lung cancer accounted for 810 new cases and 653 deaths in 2020 in Estonia, making it the 4th most prominent cancer type overall and second most prominent in males [1].

Preparations for introduction of lung cancer screening in Estonia started in 2020. The first stage in the implementation of screening was a feasibility study, conducted in 2021. Based on the results, a large-scale regional pilot study with 74 family doctors was initiated in 2022. People aged 55–74 years with an increased risk of lung cancer (according to smoking status [≥ 20 pack-years; quit < 15 years ago) or PLCOm2012$_{noRace}$ risk score (> 1.5)] were invited to the pilot study within one year through family physicians in Tartu and Tartu County, comprising 10% of the applicable population in Estonia.

During the first year, a total of 24,412 patients were evaluated; 3,708 met the inclusion criteria, of them 3,443 patients attended a LDCT scan: 6.8% of the studies had no findings, 86.1% found small, clinically insignificant findings in the lung, and 7% of patients needed either a 3-to-6-month follow-up CT scan or lung cancer investigations. In total, 31 lung cancers were diagnosed. Also, a significant stage shift in newly diagnosed cases was observed, and 60% of the cases were treated surgically.

LungFlag is a machine learning (ML) tool for calculating risk of pre-symptomatic lung cancer, that uses electronic medical record (EMR) data as input. For this study, LungFlag was used retrospectively on data collected from individuals who were referred to screening LDCT.

Objectives

1. **What is the availability and accessibility of data required for running LungFlag?**
2. **Can LungFlag be ran on the EMR data as-is?**
3. **What is the potential value of LungFlag in selecting individuals for lung cancer screening?**

Methods

LungFlag was run on EMR data of individuals who were referred to the LDCT in the Estonian regional lung cancer screening pilot study. EMR data was compiled from the pilot study database and from the Estonian national health information system (Table 1).

Aggregated data was pseudonymized and uploaded to Amazon Web Service (AWS) in Frankfurt for LungFlag calculation (Figure 1) using sFTP protocol. Calculation was initiated if at minimum information on age, sex and smoking status was available.

Top 5% of individuals at risk were selected by each method and performance was compared by AUC, ranking and average age of top 5% flagged individuals.

LungFlag was only used on retrospective dataset that was compiled by using either smoking criteria or PLCOm2012$_{noRace}$ value. Hence, performance of LungFlag was limited to a preselected dataset.

CATEGORY	PARAMETER	REQUIRED PERIODS
MINUMUM TO INITIATE LUNGFLAG CALUCLATION	Age, Sex, Smoking duration, Smoking intensity, Smoking intensity, Gessation duration, Smoking status, Leucocyte count, Platelet count	Single value or dating up to 5 years back (for blood samples)
ANTHOPOMETRY, RACE	Height, Weight, BMI, Race	Single value
BLOOD ANALYSES	Hematology and Clinical chemistry	5 years since inquiry
PREVIOUS DIAGNOSES	Previous cancers, Nicotine addiction, excessive alcohol consumption, other mental pathologies, Cardiovascular diseases, Respiratory diseases, Weight-related diseases, Abnormal radiological findings, Suspected symptoms,General risk factors	5 years since inquiry (for malignant codes entire history was requested)
SCREENING PROCESS	Initial PLCOm2012 value, Date of assessing LC risk, Date of LD-CT, compliance, LungRADs value from LD-CT, LC diagnosis from screening	Single value

Table 1: Dataset requested for each patient from either screening database or from national health database.

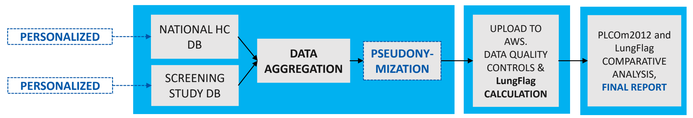

Figure 1: Data sources, data flow and pseudonymization steps for the used data.

Results

TIn total, 3,708 individuals were classified as high-risk by either smoking criteria or PLCOm20212$_{noRace}$. Out of them, 3,695 were included in the comparison (13 individuals were excluded due to previous lung cancer or missing LDCT information).

A total of 3,443 individuals underwent LDCT, and 31 cases of lung cancer were identified. Average personal smoking history was 40 years, and over 75% of the individuals were classified as current smokers (Table 2).

	Estonia	Ref (KP)
Age	53-73	45-80
Smoking Status is Current	75.5%	0.5%
Current Smoking Years – Mean	40 years	34 years
Ex Smoking Years – Mean	34.5 years	20.5 years
Quit Time – Mean	7.7 years	19.5 years
COPD (ICD code 496) ever	10%	9.3%
ICD codes 490-496* ever	17%	28%

Table 2: Estonian study population parameters. The reference data set was a random selection of population from EMR of Kaiser Permanente (KP) South Califorina, part of a study carried out with Prof. Michael Gould [2].

2,649 individuals had previous comorbidity (in the dataset: 850 had 1, 619 had 2). The top five ICD-10 diagnosis codes were COPD – J44 (10.1%), cardiac arrythmia – I49 (9.6%), atrial fibrillation and flutter – I48 (7.7%), other diseases of the respiratory system – J45 (7.6%), and heart failure – I50 (6.9%).

LungFlag scores could be calculated for all individuals, based on their EMR data.

LungFlag performed at least similarly to the PLCOm2012$_{noRace}$ model – AUC of 0.697 [95% CI 0.614–0.781] vs 0.674 [95% CI 0.594–0.759] (p=0.28), and significantly better than the smoking criterion (p< 0.01) (Fig 2).

The top 5% of individuals flagged by LungFlag were in average over 2 years younger than of those who were selected with PLCOm2012$_{noRace}$ (68.5 vs 71 years).

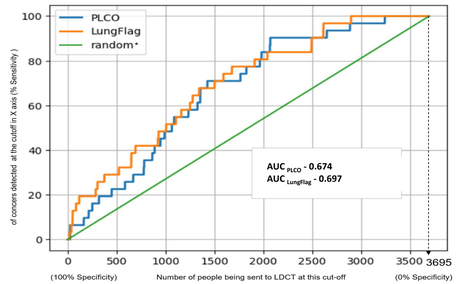

Figure 2: Comparison of AUC for PLCOm2012$_{noRace}$ and LungFlag. * Random selection reflects the proportion of cases if a random risk score is associated with each individual.

Limitations

1. **Pilot study dataset was limited to individuals preselected by PLCOm2012$_{noRace}$ or smoking criteria, but not byLungFlag (Figure 3).**
2. Missing sufficient follow-up period for flagged population.
3. Missing information (labs, spirometry, dgn codes before 2017) affected the model performance.
4. Underpowered design (total number of diagnosed lung cancers was only 31).

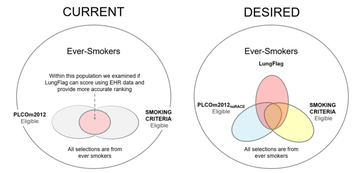

Figure 3: The desired playground for assessment on an "even floor" was to allow LungFlag to select from the same Ever-Smokers population, send the selected to LDCT and compare the results.

Next Steps

Use of LungFlag in Estonian lung cancer screening setup could be considered in the following conditions:

1. To test whether LungFlag could outperform the PLCOm2012$_{noRace}$ model when screening all ever-smokers.
2. Given availability of high-quality data, lung cancer risk assessment can be automated for individuals with pre-existing sufficient EMR data avoiding the need for repeated personal outreach.
3. Personalized risk assessment could motivate very high-risk individuals to attend initial screening and repeated assessment to come back for yearly screening.

References

1. https://gco.iarc.who.int/media/globocan/factsheets/populations/233-estonia-fact-sheet.pdf
2. Gould MK, et al. Machine Learning for Early Lung Cancer Identification Using Routine Clinical Data and Laboratory Values. Am J Respir Crit Care Med. 2021 Apr 6. DOI: 10.1164/rccm.202007-2791OC.

Tanel.Laisaar@kliinikum.ee

This study was sponsored by F. Hoffmann-La Roche Ltd