P1.104.36: LungFlag™ Risk Prediction Validation on Canadian Ever Smokers Pre-Classified as High Risk for Lung Cancer

Sukhinder Atkar-Khattra[1], Yue Jin[2], Alon Lanyado[3], Milan Obradovic[4], Abigail Walker[1], Sean Leong[1], Anne Dybuncio[1], Renelle Myers[1], Eran N Choman[3], Stephen Lam[1]

[1]British Columbia Cancer Research Institute, Vancouver Canada, [2]Data Analytics & Research, Hoffman-la Roche, Santa Clara, USA, [3]Medial EarlySign, 6 Hanagar st. hod Hasharon, Isreal, [4]Medical Affairs, Hoffman-la Roche, Laval, Canada Author 4, [5]Afliatin Author 5

Correspondence: *skhattra@bccancer.bc.ca*

IASLC 2025 World Conference on Lung Cancer
SEPTEMBER 6-9, 2025
BARCELONA, SPAIN

INTRODUCTION

- The usage of personalized risk scores is increasingly adopted to evaluate specific clinical condition susceptibility.
- Early detection of Lung Cancer (LC) is clinically important as it translates to more favorable patient outcomes.
- Thus, development of population health management approaches using risk predictors (i.e. $PLCO_{m2012}$ and EHR based LungFlag) for identification of individuals who may be in increased probability for having lung cancer is of paramount importance.

METHODS/OBJECTIVES

- LungFlag (Gould, M.K.et al., 2021. Machine learning for early lung cancer identification using **routine clinical and laboratory data**. American journal of respiratory and critical care medicine, 204(4), pp.445-453) performance was assessed on a cohort biased towards high-risk population.
- **Two independent Canadian (British Columbia) Lung Cancer screening data sets were used:**
  - (1) Lung Screening Population (LSP) – (age: 55-80 years with ≥30 years smoking history who meet either the $PLCO_{m2012}$ ≥ 1.51% or USPSTF 2013 criteria) and non-eligible population, and
  - (2) the PanCan single-arm longitudinal trial (age: 50-75 years and $PLCO_{M2008}$ ≥2% six years lung cancer risk).
- The **primary objective** is to evaluate the feasibility of LungFlag model in the detection of lung cancer in both PanCan and LSP by calculating sensitivity, specificity, positive and negative predictive values, and with area under the curve (AUC) characteristics
- To achieve the study objectives, the **LungFlag model, a machine learning algorithm, was applied to data in-order to flag the high-risk and non-high-risk patients for lung cancer. The flags were compared to the study endpoint** (actual diagnosis of lung cancer or cancer free).

DEMOGRAPHICS

Table 1: Demographic Data on Participants with Lab Data

	Total	PanCan	LSP Enrolled	LSP Not Enrolled
Participant with Lab values	3051/4851 63%	102/418 24%	2231/2305 97%	718/2128 34%
Smoking				
Former	1756	42	1180	534
Current	1295	60	1051	184
Sex				
Female	1656	46	1148	462
Male	1395	56	1083	256
Cancers				
Stage I/II	85	10	73	2
Stage III/IV	19	3	13	3
Limited	3	0	3	0
Extensive	5	0	5	0
Total Cancer:	112/156	13	94	5
Family Hx LC				
Yes	856	38	651	167
No	2195	64	1580	551
Emphysema		On LDCT	On LDCT	Self-reported
Yes	1226	74	1139	13
No	1825	28	1092	705
Education				
Less than HS	304	17	240	47
High School	694	23	546	125
Beyond HS	2053	62	1445	546
Total Number of Lab values per study	215,537	4,588	146,254	64,695

Input Values:
Outpatient lab results, comprising of blood counts, routine chemistry test results, date, test value and normal bounds | Comorbidities – medical history | Spirometry | $PLCO_{m2012}$ scores

Enrollment Criteria:
Inclusion criteria
- All Cases and controls with an existing CBC record Prediction Data Selection Approaches for Cases/Control
- PLCO Questionnaire Date: Used the same dates for both cases and control, assigning scores to patients with available information up to those dates
- Quarterly Assignment: Prediction dates were assigned once per quarter from 2015 – 2022. Imitates usage over time and allows us to take advantage of bigger cohorts and less exclusion of cases that developed cancer in later years.

Exclusion criteria
- Controls with less than 2 year of follow-up
- Cases that developed cancer within three months or more than two years after the prediction date

RESULTS

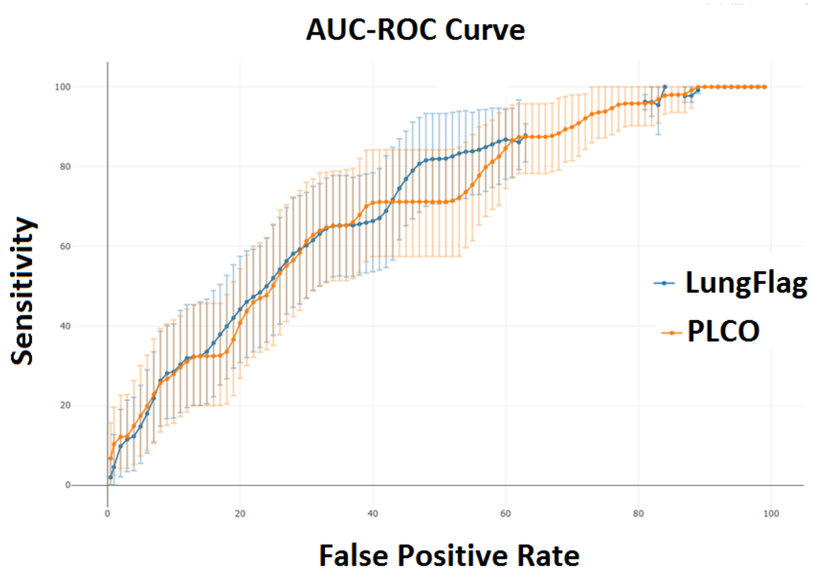

Figure 1: Comparison in AUC between PLCO and LungFlag on those who formerly or currently smoked between age 50-80 years

- The performance and ROC curve are specially based on the PLCO reference date analysis.
- Out of a total population of 3,051 individuals, **2,958 (LSP: eligible - 2,171; non-eligible-703, and PanCan: 84) met the eligibility criteria including 49 LC** (mainly due to 2 years prediction time horizon).
- The average age for cases was 69.0 with 42.9 years smoking duration compared to 64.9 and 36.5 for controls respectively.
- The proportion of participants with COPD was 35% compared to 14%, cases vs. controls, respectively.
- The **overall AUC (2-years time horizon) on those who had ever smoked between age 50-80 years was comparable between LungFlag and PLCOm2012 (0.707 vs 0.693) (Figure 1), with similar trends on the USPSTF2013 criteria (0.687 vs 0.682) subpopulation.**
- In 703 participants with 4 lung cancers who did not meet the $PLCO_{m2012}$ or USPSTF criteria, the for LungFlag AUC was 0.764 (0.434 to 0,983).

CONCLUSION

Despite being conducted among the mostly pre-selected high-risk population, LungFlag consistently demonstrates non-inferior performance compared to the PLCOm2012 or the USPSTF2013 eligibility criteria.

The model is applicable to Canadian EHR data. Using a 2-year risk horizon performed comparably well as 6 years horizons. Future studies should focus on prospective evaluation of LungFlag as an independent classifier to identify populations with an elevated risk of lung cancer for screening.

Supported by Roche, BC Cancer Foundation and Terry Fox Research Institute