HHS Public Access

Author manuscript

Ann Intern Med. Author manuscript; available in PMC 2022 June 15.

Published in final edited form as:

Ann Intern Med. 2020 November 03; 173(9): 704–713. doi:10.7326/M20-1868.

Deep learning using chest radiographs to identify high-risk smokers for lung cancer screening CT: Development and validation of a prediction model

Michael T. Lu, MD, MPH[1,2,*],

Vineet K. Raghu, PhD[1,2,*],

Thomas Mayrhofer, PhD[1,3],

Hugo J.W.L. Aerts, PhD[1,2],

Udo Hoffmann, MD, MPH[1,2]

[1]Cardiovascular Imaging Research Center (CIRC), Department of Radiology, Massachusetts General Hospital & Harvard Medical School, Boston, MA

[2]Program for Artificial Intelligence in Medicine (AIM), Brigham and Women's Hospital & Harvard Medical School, Boston, MA

[3]School of Business Studies, Stralsund University of Applied Sciences, Stralsund, Germany

Abstract

Background: Lung cancer screening with chest CT reduces lung cancer death. Centers for Medicare & Medicaid Services (CMS) eligibility criteria for lung cancer screening CT require detailed smoking information and miss many lung cancers. An automated deep learning approach based on chest radiograph (x-ray or CXR) images may identify more smokers at high-risk of lung cancer who could benefit from screening CT.

Objective: To develop and validate a convolutional neural network (CXR-LC) that predicts long-term incident lung cancer, using data commonly available in the electronic record (CXR image, age, sex, whether currently smoking).

Design: Risk prediction study.

Setting: US lung cancer screening trials.

Corresponding Author: Michael T. Lu, MD, MPH, Massachusetts General Hospital, 165 Cambridge St, Suite 400, Boston, MA 02114, P: (617) 643-0240 | F: (617) 724-4152 | mlu@mgh.harvard.edu.

*Drs. Lu and Raghu contributed equally to this article.

Reproducible Research Statement PLCO chest radiographs used for model development and validation are available by application from the National Cancer Institute (NCI). NLST chest radiographs used for external validation are available by application from the NCI and the American College of Radiology Imaging Network (ACRIN). The CXR-LC model, CXR filenames, split into development and validation datasets, and CXR-LC outputs used in this analysis are published at https://github.com/circ-ml/CXR-LC.

Participants: CXR-LC was developed in the Prostate, Lung, Colorectal, & Ovarian trial (PLCO, n=41,856). The final CXR-LC model was validated in additional PLCO smokers (n=5,615, 12-year follow-up) and National Lung Screening Trial (NLST) heavy smokers (n=5,493, 6-year follow-up). Results are reported for validation datasets only.

Measurements: Up to 12-year lung cancer incidence predicted by CXR-LC.

Results: CXR-LC had better discrimination (area under the ROC curve) for incident lung cancer than CMS eligibility (PLCO AUC 0.755 vs. 0.634, p<0.001). CXR-LC performance was similar to PLCOm2012, a state-of-the-art risk score with 11 inputs (PLCO CXR-LC AUC 0.755 vs. PLCOm2012 0.761; NLST 0.659 vs. 0.650). When compared at equal-sized screening populations, CXR-LC was more sensitive than CMS eligibility (PLCO 74.9% vs. 63.8%, p=0.012) and missed 30.7% fewer lung cancers. On decision curve analysis, CXR-LC had higher net benefit than CMS eligibility and similar benefit to PLCOm2012.

Limitation: Validation in lung cancer screening trials, and not a clinical setting.

Conclusion: CXR-LC identified smokers at high risk of incident lung cancer, beyond CMS eligibility.

Introduction

Lung cancer is the most common cause of cancer death (1). Screening high-risk smokers with chest CT reduces lung cancer mortality by 20% in the US National Lung Screening Trial (NLST) (2) and by 24% in male Dutch-Belgian Randomized Lung Cancer Screening Trial (NELSON) participants (3). In 2015, the US Centers for Medicare & Medicaid Services (CMS) determined that 55-77 year-old heavy smokers (≥30 cigarette pack-years and ≤15 years since quitting) are eligible for lung cancer screening with CT (4).

The CMS eligibility criteria were a major advance towards preventing lung cancer death, but have limitations (5, 6). Lung cancer screening eligibility criteria miss over half of lung cancers (7). Less than 5% of Americans eligible for lung cancer screening CT are screened (6, 8-10), a dismal screening rate compared to breast (64%) and colorectal (63%) cancer (11, 12). Clinicians cite barriers to broader adoption including difficulty remembering the CMS criteria, lack of time to obtain smoking pack-year and quitting history, and absent electronic medical record (EMR) reminders (13, 14). Automatically flagging smokers eligible for lung cancer screening CT in the EMR would be an important way to improve screening participation, but has proven difficult (6). Smoking pack-years are often not documented or inaccurate (15, 16), leading to underreporting of CMS screening eligibility by 54% (17). More complex risk scores like PLCOm2012 more accurately predict lung cancer, at the cost of requiring more detailed smoking and risk factor information that is often not available (Appendix Table 1) (18, 19). Lung cancer risk models that do not require pack-years or other detailed smoking information may be suited for automation in the EMR.

Our study explores a new way to identify smokers at high-risk of developing lung cancer, based on the individual's appearance on a chest radiograph (x-ray or CXR) image. We focus on chest radiography as the most common diagnostic imaging test (20). While chest radiography is ineffective for lung cancer screening (21), even "normal" radiographs

provide a window into smoking exposure and the underlying risk of lung cancer (22). We hypothesized that a deep learning convolutional neural network, a form of artificial intelligence, can learn to predict long-term incident lung cancer from patterns on chest radiograph images and other data commonly available in the EMR (age, sex, whether currently smoking). This neural network (CXR-LC) was validated in two large lung cancer screening trials. We compared CXR-LC performance to CMS eligibility and the PLCOm2012 lung cancer risk score.

Methods

Data sources

The CXR-LC model was developed and validated in the Prostate, Lung, Colorectal, and Ovarian (PLCO) cancer screening trial. PLCO enrolled men and women aged 55-74 years at 10 US sites from November 8, 1993 through July 1, 2001 (21, 23). Never, former, and current smokers were included. PLCO participants were randomized to lung cancer screening with chest radiography versus no screening. Screening arm participants had baseline (T0) and up to three annual chest radiographs (T1-T3). The primary result was that lung cancer screening with chest radiography did not reduce lung cancer mortality.

CXR-LC was externally validated in the National Lung Screening Trial (NLST) (2) of chest radiography versus low dose chest CT for lung cancer screening. NLST enrolled a community cohort of heavy smokers (≥30 pack-year history, current smoker or former smoker who quit within 15 years) aged 55-74 years old. Participants were enrolled at 33 US sites from August 2002 to April 2004; anonymized images were available from 21 sites within the American College of Radiology Imaging Network (ACRIN). The chest radiograph arm had baseline (T0) and annual (T1, T2) screening chest radiographs. The primary result was that screening CT reduced lung cancer mortality by 20% compared to chest radiography.

PLCO and NLST participants provided written informed consent for the original trials. Secondary use was approved by the National Cancer Institute, ACRIN, and our institutional review board.

Risk factors and radiographs

Baseline risk factors including age and smoking were self-reported. Participants who indicated they “smoke cigarettes regularly now” were considered current smokers (24, 25). PLCO participants stated whether they had chest radiography in the 3 years prior to enrollment, which we report as an estimate of how often existing radiographs would be available in routine clinical care.

Screening upright posterior-anterior chest radiographs were interpreted locally by centrally qualified radiologists for findings suspicious for lung cancer (e.g. lung nodule). These findings defined a “positive CXR screen” by trial criteria (2, 21, 26) and were provided to participants and their physicians. PLCO radiographs were available as scanned film; NLST radiographs available to us were digital DICOM. Chest radiograph pre-processing is detailed in the Appendix Section A.

Outcomes

The primary outcome for our study was incident lung cancer, which we defined as all lung cancers diagnosed after enrollment. The secondary outcome was lung cancer death. Lung cancer death was determined by an independent adjudication committee based on medical records including clinic notes, pathology reports and death certificates. Participants were followed until December 31, 2009, or for up to 13 (PLCO) or 8 years (NLST) (2, 21). Our validation dataset results are truncated at 12-years for PLCO and 6-years for NLST, to allow uniform comparisons between groups and to PLCOm2012 (18).

CMS eligibility and PLCOm2012

CMS lung cancer screening eligibility was defined as adults aged 55-77 with a ≥30 pack year smoking history who currently smoke or quit within the last 15 years (27). CMS criteria were based on the 2013 USPSTF recommendation, which has a slightly higher age ceiling (80 years) (28). As the maximum age in PLCO and NLST was 74 years, there is no functional difference between CMS and USPSTF criteria in our data. CMS and USPSTF criteria are the current standard of care that govern payment for lung cancer screening CT (6). PLCO included both CMS ineligible and eligible smokers. All NLST participants were CMS eligible.

We include a comparison to PLCOm2012, a validated lung cancer risk score with state-of-the-art performance (19). PLCOm2012 is a logistic regression model that predicts 6-year incident lung cancer from 11 inputs (Appendix Table 1: age, race, education, BMI, COPD, personal history of cancer, family history of lung cancer, current/former smoking status, number of cigarettes per day, years smoked, and years since smoking cessation) (18).

Development and validation datasets

52,320 PLCO chest radiograph arm participants (21) were divided into model development (80%) and validation (20%) datasets (Figure 1) (29). The development dataset was further divided 80/20 for model training (33,485/41,856) and hyperparameter tuning (8,371). The remaining PLCO participants who were current or former smokers comprised the PLCO validation dataset (5,615). The NLST external validation dataset included 5,493 participants enrolled by ACRIN who were in the chest radiography arm and had an available baseline radiograph. Training, tuning, and validation datasets were independent, meaning each individual was in only one dataset.

For the development dataset, all 85,748 available T0 and T1 radiographs were included. Training a convolutional neural network (CNN) requires a large number of training examples. A common data augmentation strategy is to train with multiple images from a single individual, which is why both T0 and T1 radiographs were included in the development dataset. For validation datasets, only the baseline T0 radiograph was considered (5,615 PLCO; 5,493 NLST) to reflect the intended use case.

Convolutional neural network

Our 12-year lung cancer incidence risk model (CXR-LC) was a fusion CNN that takes as input a single chest radiograph image along with basic information commonly available in

the EMR (age, sex, whether currently smoking) (Appendix Figure 1). The CXR-LC model is publicly available at https://github.com/circ-ml/CXR-LC.

For the imaging component of the model, we used a transfer learning approach based on our previous CXR-risk model (22) (an Inception V4 network (30) trained to predict all-cause mortality in PLCO). To avoid contamination of validation datasets, the same PLCO development dataset split used to train CXR-risk was also used to train CXR-LC. At no point were any validation dataset participants seen during development of any part of the model. Model architecture and development are described in the Appendix Section B.

We developed CXR-LC in stages—first we trained the imaging branch on smokers and nonsmokers, in order to expose the CNN to the full range of CXR appearances. We then fine-tuned the full fusion CNN in smokers only. The final CXR-LC model was validated in current or former smokers only. Like CMS eligibility and PLCOm2012, CXR-LC is not intended for use in never-smokers due to their very low lung cancer risk (31). Nevertheless, we provide results for when CXR-LC is inadvertently applied to never-smokers in Appendix Section F.

Statistical analysis

Discrimination and calibration—CXR-LC outputs a continuous probability between 0 and 1. Platt's Scaling technique (32) was used to map raw predictions to a calibrated risk score using the tuning dataset. CXR-LC was then validated in the PLCO internal validation dataset and the NLST external validation dataset. CXR-LC discrimination was assessed using the area under the receiver operating characteristic curve (AUC), with comparison to the CMS eligibility criteria and the PLCOm2012 risk model. Nested combinations of CXR-LC plus CMS eligibility and PLCOm2012 were assessed. Ninety-five percent confidence intervals and p-values for comparison of correlated ROC curves were computed using DeLong's method (33).

Calibration in the PLCO validation dataset was assessed via the Expected/Observed (E/O) ratio, defined as the number of expected incident lung cancers divided by the number of observed cancers during follow-up. Since alternate risk models predict lung cancer risk at intervals less than the PLCO and NLST follow-up, only cases in the expected timeframe were included for E/O calculations. p-values for the E/O ratio were computed using 1,000 bootstrapped samples.

Sensitivity at the same screening population size—The AUC provides an assessment of performance across all potential thresholds. To facilitate comparison at a single operating point, we compared test characteristics at a screening population size defined by CMS eligibility in the PLCO validation dataset (31, 34). "Sensitivity" and "specificity" were defined based on risk model predictions of future lung cancer, and do not refer to the sensitivity or specificity of the original chest radiograph screen. Sensitivity and specificity were compared using McNemar's test (35).

CXR-LC risk score—CXR-LC risk probabilities were converted to an ordinal CXR-LC risk score (low, indeterminate, high and very high) based on 12-year probability thresholds

(<2%, 2-<3.297%, 3.297-<8%, ≥8%). The 3.297% threshold corresponds to the CMS-defined screening population size from the previous step. The <2% and ≥8% thresholds were chosen to represent low and very high 12-year risk (Appendix Section C).

We determined the association between ordinal CXR-LC risk score and time to incident lung cancer and lung cancer death using cox proportional hazards regression. Hazard ratios were adjusted for PLCOm2012 and whether there was a positive CXR screen (18). We estimated Kaplan Meier cumulative lung cancer incidence; non-lung cancer deaths and loss to follow-up were censored. Survival curves are presented unadjusted and then adjusted for PLCOm2012 and a positive CXR screen (36).

For PLCO, 12-year predicted lung cancer incidence was compared to 12- and 6-year observed incidence. For NLST, 12-year predicted incidence was compared to up to 6-year observed incidence. Statistical analysis was performed in R and Stata. A two-sided $p<0.05$ was considered significant.

Decision curves and subgroup analysis—Decision curves evaluate the net benefit of a prediction model across a range of possible risk thresholds (37, 38). Net benefit is the weighting of the benefit of true-positives against the harms of false-positives at a particular risk-tolerance threshold (39, 40). The advantage of decision curve analysis is that it accounts for the tradeoffs between the harms and benefits of applying the model to make the decision whether to screen (39).

We present decision curves for CXR-LC, CMS eligibility, PLCOm2012, and the null scenarios of screening everyone or no one. We applied decision curve analysis to subgroups including CMS-eligible versus ineligible, positive CXR screen versus negative screen, men versus women, and by Black and White race. For these comparisons, PLCOm2012 6-year risk was adjusted to 12-year risk as detailed in the Appendix Section D.

Association with risk factors and class activation maps—We plotted pack-years versus CXR-LC probabilities using LOWESS regression. Lung cancer risk factors were compared across CXR-LC risk score categories. Gradient-weighted class activation maps (Grad-CAM) were generated from the imaging branch of CXR-LC to identify the anatomy on the chest radiograph image that contributes to predictions (41).

Results

Demographics, risk factors, and lung cancer

Table 1 presents demographics, risk factors, and lung cancer in the development and validation datasets. In PLCO development and validation datasets, 58.5% (15,883/27,164) of current or former smokers had a previous chest radiograph within the 3 years prior to enrollment. Subsequent results are reported for validation datasets only. As NLST recruited only heavy current or recent former smokers, there were more current smokers (50.4% vs. 20.2%) and higher mean pack-years (55.7 vs. 35.4) than PLCO. Despite the difference in follow-up (PLCO 12-year versus NLST 6-year), incident lung cancer was similar between

PLCO (3.7% (207/5,615)) and NLST (3.8% (206/5,493)), consistent with NLST's greater smoking burden.

CXR-LC discrimination and calibration

Discrimination for 12-year incident lung cancer was assessed using the AUC (Appendix Table 2A). The AUC describes how well a model discriminates between individuals who do and do not develop lung cancer, across the entire range of possible thresholds. An AUC of 1 is perfect performance; an AUC of 0.5 is equivalent to random chance. In the PLCO validation dataset, the CXR-LC risk score had higher AUC than CMS eligibility (0.755 vs. 0.634, $p < 0.001$) and a positive CXR screen (0.550, $p < 0.001$). CXR-LC had similar AUC to PLCOm2012 (0.761).

Among CMS-eligible smokers, CXR-LC had similar AUC to PLCOm2012 (PLCO CMS-eligible 0.681 vs. 0.679 and NLST all CMS-eligible, 0.659 vs. 0.650). Both CXR-LC and PLCOm2012 had lower AUCs in CMS-eligible smokers, consistent with observations that discrimination is more difficult in homogeneously high-risk populations (18).

Adding CXR-LC to PLCOm2012 yielded a modest improvement in AUC (+0.029 in PLCO and +0.036 in NLST, both $p<0.05$). Modest improvements are expected when adding new data to high performing models like PLCOm2012 (18). Similar trends were seen for 6-year lung cancer (Appendix Table 2B).

Model calibration was assessed using the expected/observed ratio, with ratios >1 overestimating and <1 underestimating risk. CXR-LC had good calibration in PLCO (E/O ratio 1.01 (95% CI 0.87-1.14)), similar to PLCOm2012 (1.03 (0.82-1.24)). Calibration tables are provided in Appendix Table 3.

Sensitivity at the same screening population size

Risk models were compared at risk thresholds that yielded the same screening population size as CMS eligibility, to allow a fair comparison. In the PLCO validation dataset, the CMS eligibility criteria would screen 37.9% (2,126/5,615) (Table 2A). This screening population size corresponded to a CXR-LC probability threshold of ≥3.297%. CXR-LC eligibility was more sensitive than CMS eligibility (74.9% vs. 63.8%, $p=0.012$), and had similar sensitivity to PLCOm2012 (74.4%). CXR-LC would identify 23 more of the 207 individuals who developed lung cancer (11.1%) than the CMS eligibility criteria (Table 2B). In other words, CXR-LC would miss 30.7% fewer lung cancers than CMS eligibility (23 fewer lung cancers missed out of 75 missed by CMS). Lung cancer incidence was 0.9% (22/2,547) in persons ineligible by both CXR-LC and CMS criteria, accounting for 10.6% (22/207) of all lung cancers.

CXR-LC risk score

Ordinal CXR-LC risk scores (low, indeterminate, high, very-high) were associated with incident lung cancer (Figure 2). Lung cancer incidence per 1,000 person-years for very high versus low CXR-LC risk were – PLCO: 12.4 versus 1.1; NLST: 12.7 versus 2.3 (Appendix Table 6A). Associations between CXR-LC risk and lung cancer were robust to adjustment

for a positive CXR screen and PLCOm2012 (adjusted Kaplan-Meier plot Appendix Figure 2 and adjusted hazard ratios Appendix Table 6A).

Lung cancer death

CXR-LC was developed to predict incident lung cancer; it also predicted 12-year lung cancer mortality (Appendix Figure 3). In PLCO, CXR-LC had higher AUC than CMS eligibility (0.762 vs. 0.638, p<0.001) and similar AUC to PLCOm2012 (0.762 versus 0.768, p = 0.78). Adding CXR-LC to PLCOm2012 yielded a modest improvement in AUC (+0.032 in PLCO, p=0.042 Appendix Table 2C). When compared at the same screening population size, CXR-LC had higher sensitivity than CMS eligibility (76.1% vs. 64.8%, p=0.042) and would miss 32.0% fewer lung cancer deaths (34 vs. 50) (Appendix Table 5B). Lung cancer deaths per 1,000 person-years for very high versus low CXR-LC risk were – PLCO: 9.1 versus 0.7; NLST: 8.0 versus 1.2 (Appendix Table 6B).

Decision curves and subgroup analysis

Decision curves for 12- and 6-year lung cancer in the PLCO validation dataset are provided in Figure 3. Across likely meaningful risk thresholds (approximately 1.3%-2.0% for 6-year lung cancer (5, 42); 2-5% for 12-year lung cancer), CXR-LC had similar net benefit to PLCOm2012 and higher net benefit than CMS-eligibility.

CXR-LC had similar net benefit to PLCOm2012 in subgroups divided by CMS eligibility, the result of the initial CXR screen, and by Black or White race (Appendix Figure 4). For the CMS-eligible and those with an initial positive CXR screen suspicious for lung cancer, both CXR-LC and PLCOm2012 had minimal net benefit over screening everyone in that subgroup. Most of the net benefit accrued in the CMS-ineligible.

In men, CXR-LC had higher net benefit than both PLCOm2012 and CMS eligibility. In women, CXR-LC had lower net benefit than PLCOm2012 and similar net benefit to CMS eligibility (Appendix Figure 4C). This was due to a combination of CXR-LC having better discriminatory performance in men than women (AUC 0.767 vs. 0.722), while CMS eligibility and PLCOm2012 performed worse in men than women (CMS AUC 0.618 vs. 0.663; PLCOm2012 AUC 0.735 vs. 0.806).

CXR-LC, risk factors, and heatmaps

In the PLCO validation dataset, CXR-LC risk increased with pack-year smoking history (Figure 4), both below and above 30 pack-years. Ordinal CXR-LC risk score was associated with older age, male sex, and current smoking—all expected as these were the non-image inputs into CXR-LC (Appendix Table 7). CXR-LC risk score was also associated with known lung cancer risk factors that were not inputs into the model, including pack-years, CMS screening eligibility, a positive CXR screen, and PLCOm2012 risk score. Class activation heatmaps localized predictions to the mid and upper lungs, which commonly demonstrate anatomic changes with emphysema and smoking (Appendix Figure 6).

Discussion

Chest radiographs are obtained to exclude a specific diagnosis like pneumonia. Most are "negative" in that an actionable diagnosis is not made (43). We found that a convolutional neural network (CXR-LC) can identify patterns on the image that identify smokers at high-risk of 12-year incident lung cancer and lung cancer death. CXR-LC performance was superior to CMS eligibility criteria, the current US standard, missing 30.7% fewer people with incident lung cancer at the same screening population size. CXR-LC discrimination (AUC) and net benefit (decision curves) were similar to a state-of-the-art risk score (PLCOm2012), without requiring pack-years or other in-depth smoking information. CXR-LC risk predictions were robust to adjustment for whether the screening CXR was "positive" for lung cancer and the PLCOm2012 risk score. CXR-LC's ability to risk stratify CMS eligible smokers was externally validated in NLST for 6-year lung cancer. CXR-LC was associated with smoking pack-years, and class activation heatmaps localized predictions to the mid and upper lungs. This supports the hypothesis that CXR-LC may identify radiographic changes related to smoking.

A major issue in lung cancer screening is that <5% of US CMS screening-eligible smokers are screened (6, 8-10). Our goal with CXR-LC was a pragmatic approach to identify high risk smokers when there is not sufficient smoking information to apply CMS or other risk scores (44, 45). We designed CXR-LC to take as input commonly available data (chest radiograph, age, sex, current smoking status). In most modern health systems, chest radiograph images are archived electronically to a PACS (Picture Archiving and Communication System) (46). PACS exists in parallel with the EMR. While some EMRs may not allow viewing of CXR images, these images are in fact available through PACS. Age and sex are embedded in the header metadata of all modern chest radiograph images. Whether there is current smoking adds information about future exposure that we felt would be difficult to extract from the image. In the US, documentation of current smoking status is mandated by the CMS Meaningful Use regulation and often available (47). 2013 data from 436,652 persons from 15 US community health centers found that 90% had a documented smoking status (current/former/never) in the EMR (47). A US Veterans Administration lung screening program found that smoking status was available in 99% of primary care patients aged 55-80 in 2013-2015, while pack-years could not be calculated in 39% (36,555/93,033) (16). The UK Quality and Outcomes Framework (QOF) incentivizes documentation of smoking status, with 87% (849,824/982,217) aged ≥55 years having a documented smoking status in 2006/2007 (48). If current smoking status is not available, CXR-LC results for both former and current smokers could be provided.

While the cost ($21 in US CMS (49); £25 in UK NHS (50)) and radiation dose (equivalent to 10 days natural background radiation (51)) is low, we would not recommend chest radiography solely to assess lung cancer risk. Instead, a pragmatic future implementation of CXR-LC could analyze existing outpatient smokers' chest radiographs using an automated EMR tool. Vendor and free open-source tools are available to integrate automatic AI algorithms like CXR-LC into PACS (52). The process of analyzing an image using CXR-LC takes less than half a second using standard chest radiographs on a local consumer-grade computer. High CXR-LC risk would trigger an EMR alert to perform a targeted interview

to assess risk and discuss lung cancer screening. CXR-LC could then inform the shared decision-making process, as a complement to CMS eligibility and PLCOm2012. This two-stage approach using CXR-LC as an automated EMR screen followed by a targeted risk and screening eligibility interview addresses the primary goal of getting more high-risk smokers into the screening pipeline, while retaining the confidence of traditional risk scores. For persons eligible by CXR-LC but not CMS criteria, we note that the 12-year lung cancer incidence of 5.6% (53/942) was only slightly lower than the 6.2% (132/2,126) in the entire CMS-eligible group (Table 2B). In this case application of another risk score (e.g. PLCOm2012) may aid decision-making. While a convolutional neural network is inherently more complex than binary eligibility criteria (CMS) or a regression model (PLCOm2012), once automated in the EMR it has the potential to consume less time from the physician's perspective.

Screening based on individual risk is increasingly accepted, though there is not consensus on a threshold (5). The National Comprehensive Cancer Network practice guidelines suggest screening smokers who are CMS eligible or have a PLCOm2012 6-year risk ≥1.3% (42). Interim results from the International Lung Screening Trial (ILST) demonstrate that screening based on a PLCOm2012 6-year risk ≥1.5% was more sensitive than the USPSTF criteria (53). The Cancer Care Ontario Screening Pilot for People at High Risk is a program evaluation of lung screening CT in smokers with a PLCOm2012 6-year risk ≥2% (5).

In our study, adding CXR-LC to PLCOm2012 yielded a modest improvement in AUC. For 12-year lung cancer, the AUC improved +0.029 to 0.790 in PLCO and +0.047 to 0.823 in the CMS-ineligible subgroup (Appendix Table 2A). This suggests that combining CXR-LC and risk models like PLCOm2012 could further improve risk prediction, when both chest radiograph images and detailed lung cancer risk factors are available.

To our knowledge, CXR-LC is the first risk score to predict long-term incident lung cancer based on a chest radiograph image. CXR-LC evaluates the raw image using a CNN, without human input as to what parts of the image have value. This is in contrast to efforts based on the radiologist's interpretation of chest CT (54). The PLCO2012results model incorporates the radiologists' interpretation for lung nodule size, type, and growth on 3 annual lung screening CTs to predict incident lung cancer in NLST (55). Radiologist findings of emphysema on CT are associated with prevalent and incident lung cancer in NLST and International Early Lung Cancer Action Program (I-ELCAP) (56, 57). These CT-based approaches require one or more completed screening CTs, and thus do not apply to the arguably more important population that has never been screened. Many of these older heavy smokers will have existing chest radiographs that could be repurposed for CXR-LC.

As an estimate of how often existing radiographs are available, we found that 58.5% of PLCO smokers had a chest radiograph in the 3 years before enrollment. PLCO enrolled participants between 1993-2001. For a better estimate in today's general outpatient population, we searched our hospital's Research Patient Database Registry (RPDR) (58) for smokers aged 55-74 years who had an outpatient general medical visit (e.g. annual exam) in 2019, finding that 40.0% (5,635/14,097) had a chest radiograph from 2016-2019. This suggests that many older smokers will have recent chest radiographs.

Limitations of this study should be considered. First, PLCO and NLST included community dwelling adults aged 55-74 who had screening frontal posterior-anterior chest radiographs. Most participants (PLCO 87%, NLST 93%) were non-Hispanic Caucasians. Smoking is declining in the US. Whether our results generalize to other demographics remains to be seen (59). Second, CXR-LC was validated using radiographs from lung cancer screening trials, not radiographs obtained for symptoms. Patients with lung cancer symptoms are not eligible for lung cancer screening CT – they should instead have diagnostic chest CT or another test (4). Third, a central assumption of risk-based screening is that the highest risk will derive the most benefit. However we did not account for the competing risk of non-lung cancer morbidity (e.g. myocardial infarction) or potential harms (overdiagnosis) (60). Persons with high CXR-LC risk tended to be older, and the threshold where benefits outweigh costs after considering span and quality of life should be evaluated. Fourth, CXR-LC uses low resolution (224 pixel) chest radiograph images. While our evaluation of performance considers the time to event, our loss function (method used to train the model) used only a binary outcome (event/no event) during model training. Future models using higher resolution images and that incorporate the time to event during training may improve performance (61, 62). Fifth, barriers to lung cancer screening like poor awareness and buy-in may not be addressed by prediction tools (6). EMR integration is complex and requires quality assurance. Whether CXR-LC improves lung cancer screening participation and effectiveness will need to be assessed in clinical trials.

In conclusion, a convolutional neural network can find patterns on chest radiography that identify smokers at high long-term risk of lung cancer, beyond CMS criteria for lung cancer screening eligibility. High risk individuals may benefit from lung cancer screening CT.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the study participants, investigators, and the National Cancer Institute for data collected in the PLCO and NLST trials. Original data collection for the ACRIN 6654 trial (NLST) was supported by NCI Cancer Imaging Program grants. The statements contained herein are solely those of the authors and do not represent or imply concurrence or endorsements by the above organizations.

Conflict of Interest Disclosures

A graphics processing unit used for this research was donated to Dr. Lu as an unrestricted gift through the Nvidia Corporation Academic Program. Dr. Lu reported research funding as a co-investigator to MGH from Kowa Company Limited and Medimmune/Astrazeneca and receiving personal fees from PQBypass unrelated to this work. Dr. Lu has common stock in Nvidia and AMD. Dr. Raghu has common stock in Nvidia, Alphabet, and Apple. Dr. Raghu was supported by NIH/NHLBI T32 HL076136. Dr. Aerts reported receiving personal fees from Sphera, Genospace, and Onc.AI outside the submitted work. Dr. Hoffmann reported receiving research support on behalf of his institution from Duke University (Abbott), HeartFlow, Kowa Company Limited, and MedImmune/Astrazeneca; and receiving consulting fees from Duke University (NIH), and Recor Medical unrelated to this research. No other potential conflict of interest relevant to this article was reported.

References

1. National Cancer Institute. Cancer Stat Facts: Common Cancer Sites. Accessed at https://seer.cancer.gov/statfacts/html//common.html on May 2, 2019.

2. National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011;365(5):395–409. [PubMed: 21714641]

3. de Koning HJ, van der Aalst CM, de Jong PA, Scholten ET, Nackaerts K, Heuvelmans MA, et al. Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial. N Engl J Med. 2020;372(6):503–13.

4. Decision Memo for Screening for Lung Cancer with Low Dose Computed Tomography (LDCT) (CAG-00439N). 2015.

5. Tammemagi MC. Selecting lung cancer screenees using risk prediction models-where do we go from here. Transl Lung Cancer Res. 2018;7(3):243–53. [PubMed: 30050763]

6. Wang GX, Baggett TP, Pandharipande PV, Park ER, Percac-Lima S, Shepard JO, et al. Barriers to Lung Cancer Screening Engagement from the Patient and Provider Perspective. Radiology. 2019;290(2):278–87. [PubMed: 30620258]

7. Wang Y, Midthun DE, Wampfler JA, Deng B, Stoddard SM, Zhang S, et al. Trends in the proportion of patients with lung cancer meeting screening criteria. JAMA. 2015;313(8):853–5. [PubMed: 25710663]

8. Jemal A, Fedewa SA. Lung Cancer Screening With Low-Dose Computed Tomography in the United States-2010 to 2015. JAMA Oncol. 2017;3(9):1278–81. [PubMed: 28152136]

9. Richards TB, Doria-Rose VP, Soman A, Klabunde CN, Caraballo RS, Gray SC, et al. Lung Cancer Screening Inconsistent With U.S. Preventive Services Task Force Recommendations. Am J Prev Med. 2019;56(1):66–73. [PubMed: 30467092]

10. Smith RA, Andrews KS, Brooks D, Fedewa SA, Manassaram-Baptiste D, Saslow D, et al. Cancer screening in the United States, 2019: A review of current American Cancer Society guidelines and current issues in cancer screening. CA Cancer J Clin. 2019;69(3):184–210. [PubMed: 30875085]

11. American Cancer Society. Breast Cancer Facts & Figures 2017-2018. American Cancer Society. 2018. Accessed at https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2017-2018.pdf on May 5, 2019.

12. American Cancer Society. Colorectal Cancer Facts & Figures 2017-2019. American Cancer Society. 2019. Accessed at https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/colorectal-cancer-facts-and-figures/colorectal-cancer-facts-and-figures-2017-2019.pdf on May 5, 2019.

13. Melzer AC, Golden SE, Ono SS, Datta S, Triplette M, Slatore CG. "We Just Never Have Enough Time": Clinician Views of Lung Cancer Screening Processes and Implementation. Ann Am Thorac Soc. 2020.

14. Triplette M, Kross EK, Mann BA, Elmore JG, Slatore CG, Shahrir S, et al. An Assessment of Primary Care and Pulmonary Provider Perspectives on Lung Cancer Screening. Ann Am Thorac Soc. 2018;15(1):69–75. [PubMed: 28933940]

15. Detterbeck FC, Mazzone PJ, Naidich DP, Bach PB. Screening for lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest. 2013;143(5 Suppl):e78S–e92S. [PubMed: 23649455]

16. Kinsinger LS, Anderson C, Kim J, Larson M, Chan SH, King HA, et al. Implementation of Lung Cancer Screening in the Veterans Health Administration. JAMA Intern Med. 2017;177(3):399–406. [PubMed: 28135352]

17. Modin HE, Fathi JT, Gilbert CR, Wilshire CL, Wilson AK, Aye RW, et al. Pack-Year Cigarette Smoking History for Determination of Lung Cancer Screening Eligibility. Comparison of the Electronic Medical Record versus a Shared Decision-making Conversation. Ann Am Thorac Soc. 2017;14(8):1320–5. [PubMed: 28406708]

18. Tammemagi MC, Katki HA, Hocking WG, Church TR, Caporaso N, Kvale PA, et al. Selection criteria for lung-cancer screening. N Engl J Med. 2013;368(8):728–36. [PubMed: 23425165]

19. Katki HA, Kovalchik SA, Petito LC, Cheung LC, Jacobs E, Jemal A, et al. Implications of Nine Risk Prediction Models for Selecting Ever-Smokers for Computed Tomography Lung Cancer Screening. Ann Intern Med. 2018;169(1):10–9. [PubMed: 29800127]

20. Ron E Cancer risks from medical radiation. Health Phys. 2003;85(1):47–59. [PubMed: 12852471]

21. Oken MM, Hocking WG, Kvale PA, Andriole GL, Buys SS, Church TR, et al. Screening by chest radiograph and lung cancer mortality: the Prostate, Lung, Colorectal, and Ovarian (PLCO) randomized trial. JAMA. 2011;306(17):1865–73. [PubMed: 22031728]
22. Lu MT, Ivanov A, Mayrhofer T, Hosny A, Aerts H, Hoffmann U. Deep Learning to Assess Long-term Mortality From Chest Radiographs. JAMA Netw Open. 2019;2(7):e197416. [PubMed: 31322692]
23. Prorok PC, Andriole GL, Bresalier RS, Buys SS, Chia D, Crawford ED, et al. Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. Control Clin Trials. 2000;21(6 Suppl):273S–309S. [PubMed: 11189684]
24. Prostate Lung Colorectal and Ovarian Cancer Screening Trial. Baseline Questionnaire for Male Participants (BQM3). 1999. Accessed at https://cdas.cancer.gov/static/cdas/documents/studies/plco/forms/BQM3.4c5f1fbf22b8.pdf on May 25, 2020.
25. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011;365(5):395–409. [PubMed: 21714641]
26. Tammemagi MC, Freedman MT, Pinsky PF, Oken MM, Hu P, Riley TL, et al. Prediction of true positive lung cancers in individuals with abnormal suspicious chest radiographs: a prostate, lung, colorectal, and ovarian cancer screening trial study. J Thorac Oncol. 2009;4(6):710–21. [PubMed: 19404219]
27. Chin J, Syrek Jensen T, Ashby L, Hermansen J, Hutter JD, Conway PH. Screening for lung cancer with low-dose CT--translating science into Medicare coverage policy. N Engl J Med. 2015;372(22):2083–5. [PubMed: 26017822]
28. Moyer VA. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2014;160(5):330–8. [PubMed: 24378917]
29. Liu Y, Chen PC, Krause J, Peng L. How to Read Articles That Use Machine Learning: Users' Guides to the Medical Literature. JAMA. 2019;322(18):1806–16. [PubMed: 31714992]
30. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. 2016. Accessed at https://arxiv.org/abs/1602.07261 on April 4, 2018.
31. Tammemagi MC, Church TR, Hocking WG, Silvestri GA, Kvale PA, Riley TL, et al. Evaluation of the lung cancer risks at which to screen ever- and never-smokers: screening rules applied to the PLCO and NLST cohorts. PLoS Med. 2014;11(12):e1001764. [PubMed: 25460915]
32. Platt J Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. Advances in Large Margin Classifiers. 1999;10(3):61–74.
33. Sun X, Xu W. Fast implementation of DeLong's algorithm for comparing the areas under correlated receiver operating characteristic curves. IEEE Signal Processing Letters. 2014;21(11):1389–93.
34. Ten Haaf K, Jeon J, Tammemagi MC, Han SS, Kong CY, Plevritis SK, et al. Risk prediction models for selection of lung cancer screening candidates: A retrospective validation study. PLoS Med. 2017;14(4):e1002277. [PubMed: 28376113]
35. McNemar Q Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika. 1947;12(2):153–7. [PubMed: 20254758]
36. Therneau TM, Croson CS, Atkinson EJ. Adjusted survival curves. R-project. 2015. Accessed at https://cran.r-project.org/web/packages/survival/vignettes/adjcurve.pdf on June 21, 2020.
37. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. BMJ. 2016;352:i6. [PubMed: 26810254]
38. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. Diagn Progn Res. 2019;3:18. [PubMed: 31592444]
39. Raji OY, Duffy SW, Agbaje OF, Baker SG, Christiani DC, Cassidy A, et al. Predictive accuracy of the Liverpool Lung Project risk model for stratifying patients for computed tomography screening for lung cancer: a case-control and cohort validation study. Ann Intern Med. 2012;157(4):242–50. [PubMed: 22910935]
40. Localio AR, Goodman S. Beyond the usual prediction accuracy metrics: reporting results for clinical decision making. Ann Intern Med. 2012;157(4):294–5. [PubMed: 22910942]

41. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. arXiv e-prints; 2016.
42. Wood DE, Kazerooni EA, Baum SL, Eapen GA, Ettinger DS, Hou L, et al. Lung Cancer Screening, Version 3.2018, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw. 2018;16(4):412–41. [PubMed: 29632061]
43. Tape TG, Mushlin AI. The utility of routine chest radiographs. Ann Intern Med. 1986;104(5):663–70. [PubMed: 3516043]
44. Polubriaginof F, Salmasian H, Albert DA, Vawdrey DK. Challenges with Collecting Smoking Status in Electronic Health Records. AMIA Annu Symp Proc. 2017;2017:1392–400. [PubMed: 29854208]
45. Uzuner O, Goldstein I, Luo Y, Kohane I. Identifying patient smoking status from medical discharge records. J Am Med Inform Assoc. 2008;15(1):14–24. [PubMed: 17947624]
46. Dikici E, Bigelow M, Prevedello LM, White RD, Erdal BS. Integrating AI into radiology workflow: levels of research, production, and feedback maturity. J Med Imaging (Bellingham). 2020;7(1):016502. [PubMed: 32064302]
47. Fortmann SP, Bailey SR, Brooks NB, Hitsman B, Rittner SS, Gillespie SE, et al. Trends in smoking documentation rates in safety net clinics. Health Serv Res. 2020;55(2):170–7. [PubMed: 31930738]
48. Simpson CR, Hippisley-Cox J, Sheikh A. Trends in the epidemiology of smoking recorded in UK general practice. Br J Gen Pract. 2010;60(572):e121–7. [PubMed: 20202355]
49. American College of Radiology. Medicare Physician Fee Schedule Impacts for 70,000-series CPT codes. American College of Radiology. 2019. Accessed at https://www.acr.org/-/media/ACR/Files/Advocacy/MPFS/2019-Impact-Table-70000.pdf?la=en on November 6, 2019.
50. GoCompare. The Bill of Health. 2018. Accessed at https://www.gocompare.com/health-insurance/the-bill-of-health/ on January 9, 2020.
51. radiologyinfo.org. Radiation dose to adults from common imaging examinations. Radiological Society of North America and American College of Radiology. 2018. Accessed at https://www.acr.org/-/media/ACR/Files/Radiology-Safety/Radiation-Safety/Dose-Reference-Card.pdf on November 6, 2019.
52. Sohn JH, Chillakuru YR, Lee S, Lee AY, Kelil T, Hess CP, et al. An Open-Source, Vender Agnostic Hardware and Software Pipeline for Integration of Artificial Intelligence in Radiology Workflow. J Digit Imaging. 2020.
53. Lam S, Myers R, Ruparel M, Atkar-Khattra SK, Stone E, Manser R, et al. Lung Cancer Screenee Selection by USPSTF Versus PLCOm2012 Criteria – Interim ILST Findings. J Thorac Oncol. 2019;14(10):S4–S5.
54. O'Dowd EL, Ten Haaf K. Lung cancer screening: enhancing risk stratification and minimising harms by incorporating information from screening results. Thorax. 2019;74(9):825–7. [PubMed: 31363021]
55. Tammemagi MC, Ten Haaf K, Toumazis I, Kong CY, Han SS, Jeon J, et al. Development and Validation of a Multivariable Lung Cancer Risk Prediction Model That Includes Low-Dose Computed Tomography Screening Results: A Secondary Analysis of Data From the National Lung Screening Trial. JAMA Netw Open. 2019;2(3):e190204. [PubMed: 30821827]
56. Henschke CI, Yip R, Boffetta P, Markowitz S, Miller A, Hanaoka T, et al. CT screening for lung cancer: Importance of emphysema for never smokers and smokers. Lung Cancer. 2015;88(1):42–7. [PubMed: 25698134]
57. Yong PC, Sigel K, de-Torres JP, Mhango G, Kale M, Kong CY, et al. The effect of radiographic emphysema in assessing lung cancer risk. Thorax. 2019.
58. Nalichowski R, Keogh D, Chueh HC, Murphy SN. Calculating the benefits of a Research Patient Data Repository. AMIA Annu Symp Proc. 2006:1044. [PubMed: 17238663]
59. Aldrich MC, Mercaldo SF, Sandler KL, Blot WJ, Grogan EL, Blume JD. Evaluation of USPSTF Lung Cancer Screening Guidelines Among African American Adult Smokers. JAMA Oncol. 2019.
60. Lu MT, Onuma OK, Massaro JM, D'Agostino RB Sr, O'Donnell CJ, Hoffmann U. Lung Cancer Screening Eligibility in the Community: Cardiovascular Risk Factors, Coronary Artery Calcification, and Cardiovascular Events. Circulation. 2016;134(12):897–9. [PubMed: 27647299]

61. Baltruschat IM, Nickisch H, Grass M, Knopp T, Saalbach A. Comparison of Deep Learning Approaches for Multi-Label Chest X-Ray Classification. Sci Rep. 2019;9(1):6381. [PubMed: 31011155]

62. Avati A, Duan T, Jung K, Shah NH, Ng A. Countdown Regression: Sharp and Calibrated Survival Predictions. arXiv e-prints; 2018.

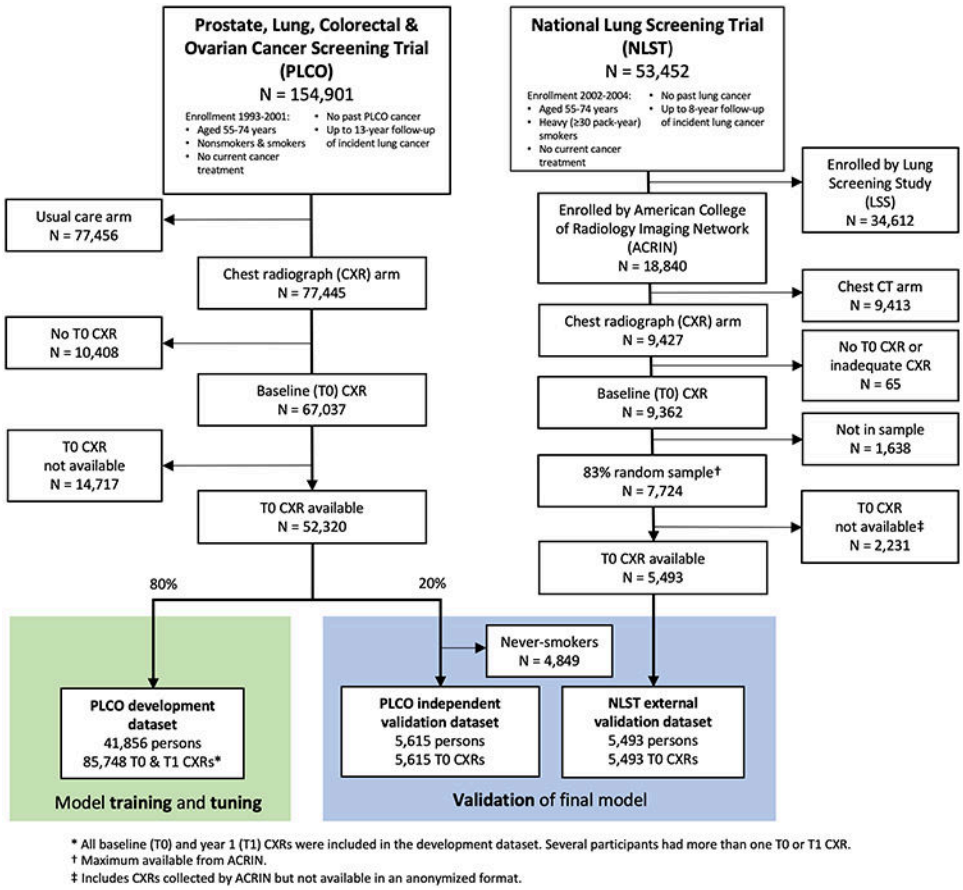

Figure 1: Datasets for CXR-LC model development and validation.

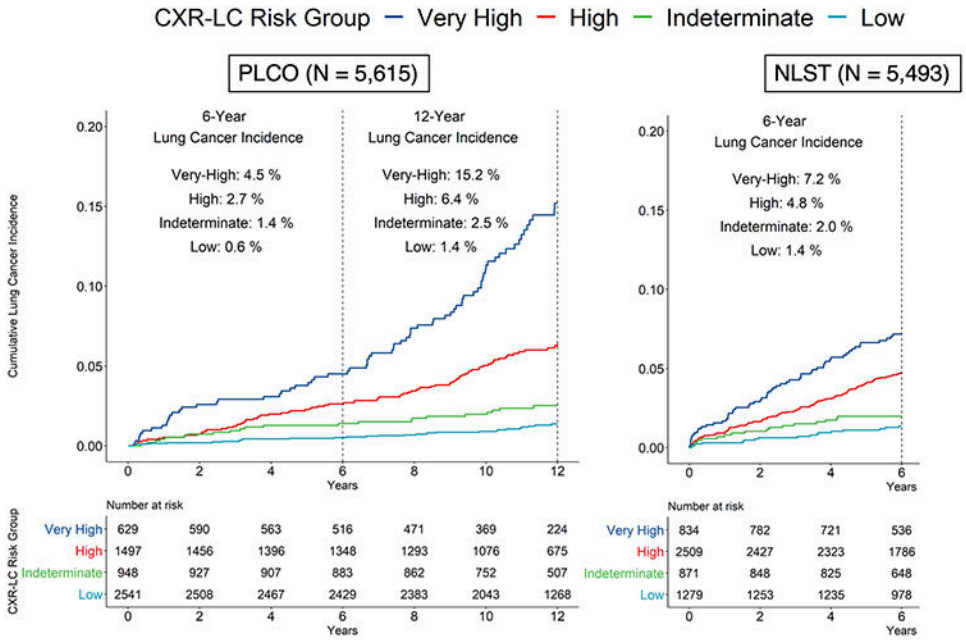

Figure 2: Inverted Kaplan-Meier plot of incident lung cancer by CXR-LC risk score in validation datasets.

Data are presented without adjustment; Kaplan Meier plots adjusted for PLCOm2012 and a positive CXR screen yielded similar results (Appendix Figure 2).

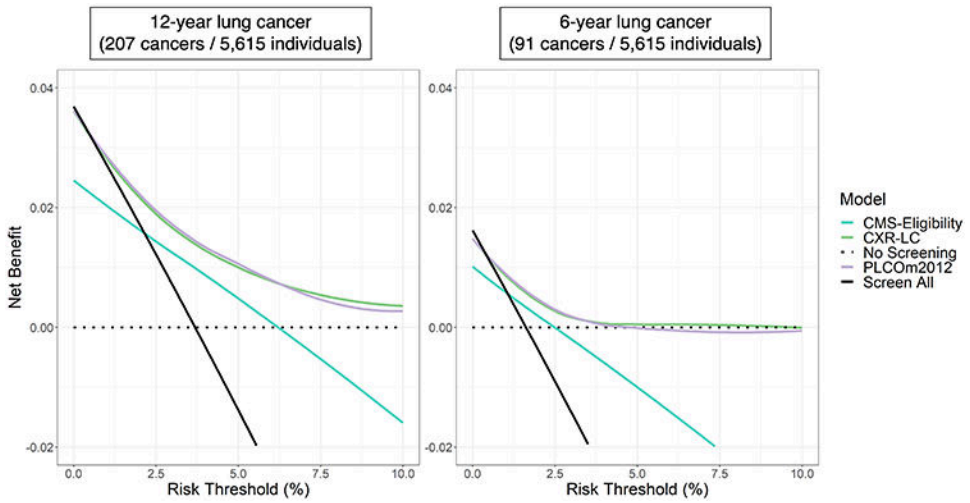

Figure 3: Decision curves for 12- and 6-year incident lung cancer in the PLCO validation dataset.

Decision curve analysis weights the harms against the benefits of using a prediction model to make clinical decisions, across a range of risk thresholds. CXR-LC had higher net benefit than CMS eligibility and similar net benefit to PLCOm2012.

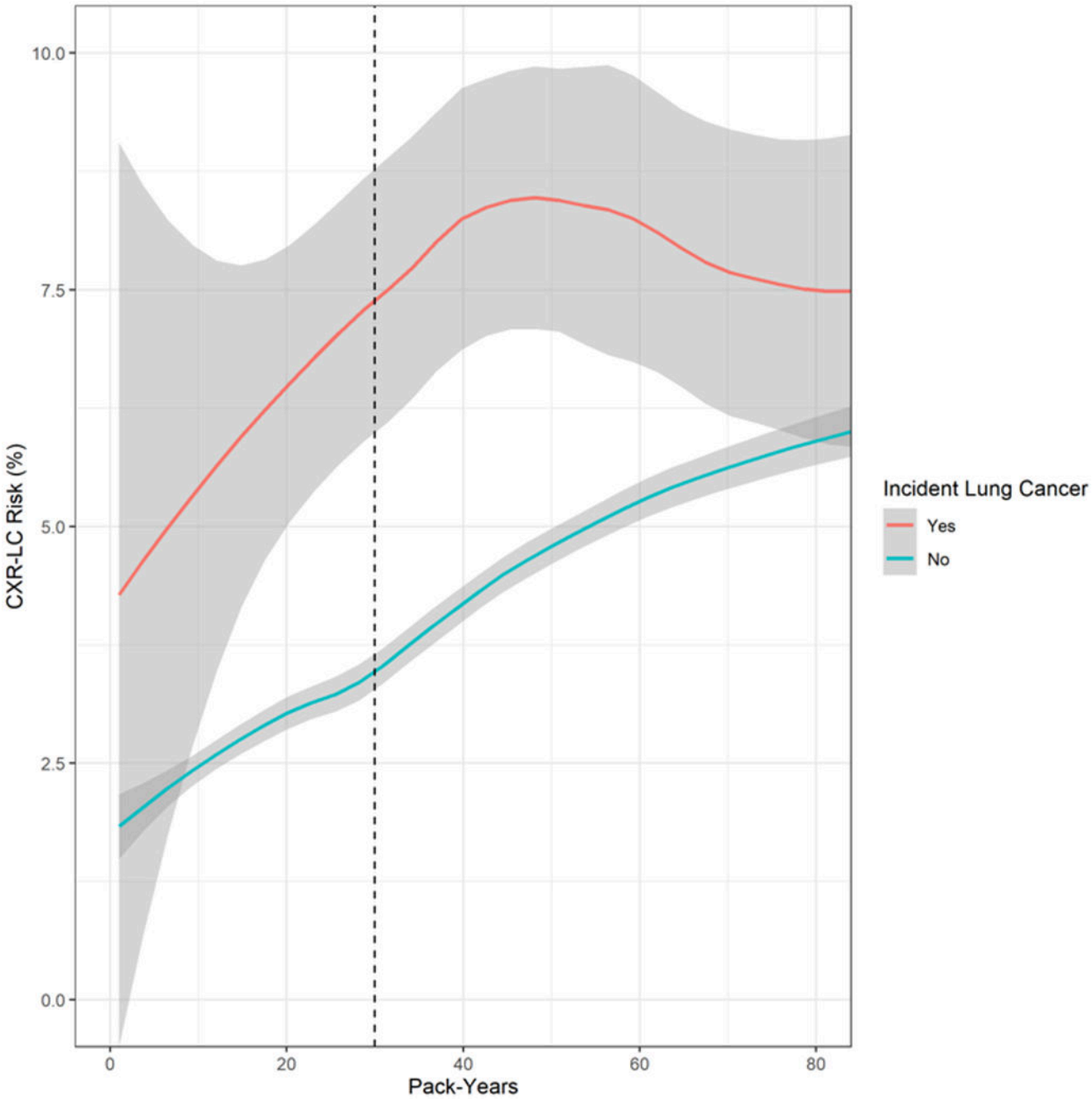

Figure 4: Pack-year smoking history versus CXR-LC risk probability in the PLCO validation dataset.

CXR-LC risk probabilities were strongly associated with pack-years, both below and above the 30 pack-year CMS eligibility threshold (vertical dashed line). At pack-year levels with sufficient samples, CXR-LC predicted significantly higher risk in those who developed lung cancer (red line) versus those who were lung cancer-free (teal line) after 12 years. Gray shading denotes 95% confidence interval.

Table 1:
Baseline characteristics and incident lung cancer.

The CXR-LC convolutional neural network was first developed in smokers and non-smokers, then fine-tuned in the subset of smokers. The final CXR-LC network was validated in smokers from the PLCO independent validation and NLST external validation datasets.

Characteristic	PLCO			NLST
	Development (Smokers and non-smokers, N = 41,856)	Development (Smokers only, N = 22,711)	Independent Validation (Smokers only, N = 5,615)	External Validation (Smokers only, N = 5,493)
# of screening radiographs *	85,478	46,829	5,615	5,493
Radiograph in 3 years prior to enrollment	22,390 / 39,970 (56.0%)	12,703 / 21,793 (58.3%)	3,180 / 5,371 (59.2%)	NA
Age, mean (SD)	62.4 (5.4)	62.2 (5.3)	62.1 (5.3)	61.7 (5.0)
Sex (male)	21,648 / 41,856 (51.7%)	13,660 / 22,711 (60.1%)	3,373 / 5,615 (60.1%)	3,037 / 5,493 (55.3%)
Race/Ethnicity				
Caucasian, Non-Hispanic	36,295 / 41,856 (86.7%)	19, 797 / 22,711 (87.2%)	4,889 / 5,615 (87.1%)	5,105 / 5,493 (92.9%)
African American, Non-Hispanic	2,451 / 41,856 (5.9%)	1,411 / 22,711 (6.2%)	356 / 5,615 (6.3%)	221 / 5,493 (4.0%)
Hispanic	775 / 41,856 (1.9%)	460 / 22,711 (2.0%)	117 / 5,615 (2.1%)	167 / 5,493 (3.0%)
Asian/Other	2,335 / 41,856 (5.6%)	1,043 / 22,711 (4.6%)	253 / 5,615 (4.5%)	0 / 5,493 (0.0%)
BMI				
Underweight (<18.5 kg/m2)	281 / 41,275 (0.7%)	135 / 22, 443 (0.6%)	41 / 5,547 (0.7%)	45 / 5,484 (0.8%)
Obese (≥30 kg/m2)	9,978 / 41,275 (24.2%)	5,544 / 22,443 (24.7%)	1,387 / 5,547 (25%)	1,518 / 5,484 (27.7%)
Smoking status				
Never	19,145 / 41,856 (45.7%)		-	-
Current	4,392 / 41,856 (10.5%)	4,392 / 22,711 (19.3%)	1,133 / 5,615 (20.2%)	2,769 / 5,493 (50.4%)
Former	18,319 / 41,856 (44.9%)	18,319 / 22,711 (80.7%)	4,482 / 5,615 (79.8%)	2,724 / 5,493 (49.6%)
Years since cessation, mean (SD) †	20.4 (11.9)	20.4 (11.9)	16.2 (13.4)	7.2 (4.9)
Pack-years, mean (SD)	19.3 (27.7)	35.2 (29.0)	35.4 (29.0)	55.7 (23.5)
CMS lung cancer screening eligibility	8,456 / 41,856 (20.2%)	8,456 / 22,711 (37.2%)	2,126 / 5,615 (37.9%)	-
Positive CXR screen ‡	3,770 / 41,856 (9.0%)	2,208 / 22,711 (9.7%)	573 / 5,615 (10.2%)	505 / 5,493 (9.2%)
Lung cancer incidence §	962 / 41,856 (2.3%)	880 / 22,711 (3.9%)	207 / 5,615 (3.7%)	206 / 5,493 (3.8%)
Lung cancer mortality §	634 / 41,856 (1.5%)	586 / 22,711 (2.6%)	142 / 5,615 (2.5%)	113 / 5,493 (2.1%)

Abbreviations: PLCO, Prostate, Lung, Colorectal, Ovarian Cancer Screening Trial; NLST, National Lung Screening Trial; SD, Standard Deviation; BMI, Body Mass Index; CXR, Chest X-ray

*The development dataset included both enrollment (T0) and the first annual (T1) chest radiographs; validation dataset included a single T0 radiograph per participant.

†Computed only in former smokers.

‡Radiographs were interpreted by centrally qualified radiologists for findings suspicious for lung cancer (e.g. lung nodule).

§Lung cancer incidence and death are reported at 12-years for PLCO and 6-years for NLST validation datasets

Table 2:

A) Test characteristics for 12-year lung cancer. CXR-LC had higher sensitivity than CMS eligibility (p=0.012) and similar sensitivity to PLCOm2012.

Risk Model	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Number Screened (%)	Lung Cancers Included (%)
CMS eligibility *	63.8% (57.2, 70.3)	63.1% (61.8, 64.4)	6.2% (5.2, 7.2)	97.9% (97.4, 98.3)	2126 (37.9%)	132 (63.8%)
CXR-LC †	74.9% (69.0, 80.8)	63.6% (62.3, 64.8)	7.3% (6.2, 8.4)	98.5% (98.1, 98.9)	2126 (37.9%)	155 (74.9%)
PLCOm2012 ‡	74.4% (68.5, 80.3)	63.5% (62.3, 64.8)	7.2% (6.1, 8.3)	98.4% (98.1, 98.9)	2126 (37.9%)	154 (74.4%)

B) Percent with 12-year incident lung cancer, by CXR-LC and CMS eligibility.			
	CMS eligible	CMS ineligible	Total
CXR-LC eligible	8.6% (102/1,184)	5.6% (53/942)	7.3% (155/2,126)
CXR-LC ineligible	3.2% (30/942)	0.9% (22/2,547)	1.5% (52/3,489)
Total	6.2% (132/2,126)	2.1% (75/3,489)	3.7% (207/5,615)

C) Percent with 12-year incident lung cancer, by CXR-LC and PLCOm2012 eligibility.			
	PLCOm2012 Eligible	PLCOm2012 Ineligible	Total
CXR-LC eligible	9.0% (122/1,352)	4.3% (33/774)	7.3% (155/2,126)
CXR-LC ineligible	4.1% (32/774)	0.7% (20/2,715)	1.5% (52/3,489)
Total	7.2% (154/2, 126)	1.5% (53/3,489)	3.7% (207/5,615)

Test characteristics are reported for the PLCO validation dataset. Comparisons are made at risk thresholds that yield an equally-sized screening population (number screened) to the CMS eligibility criteria.

* CMS eligibility defined as current or recent former (quit within 15 years) heavy smokers (≥30 pack-years) aged 55-77

† CXR-LC convolutional neural network based on the chest radiograph image, age, sex, and current vs. former smoking status.

‡ PLCOm2012 regression model based on age, race, education, body mass index, chronic obstructive pulmonary disease, history of cancer, family history of lung cancer, current vs. former smoking status, number of cigarettes per day, years smoked, and years since smoking cessation.

Abbreviations: PPV, Positive Predictive Value; NPV, Negative Predictive Value; CI: Confidence Interval; CMS: Centers for Medicare and Medicaid Services