

Lung Cancer Risk Using Never Smokers' Chest X-Rays: Validation of a Deep Learning-based Model

PURPOSE

Lung cancer is the most common cause of cancer death. In the United States, 10-20% of lung cancers occur in “never-smokers” - those who never smoked cigarettes or smoked fewer than 100 cigarettes in their lifetime. Centers for Medicare and Medicaid Services (CMS) lung cancer screening criteria do not recommend screening never-smokers; however, never-smokers often present with more advanced lung cancer than those who smoke. In this study, we tested whether a deep learning model CXR-Lung-Risk could identify never-smokers at high risk for lung cancer using chest x-rays (CXRs) from the electronic medical record.

METHODS AND MATERIALS

The CXR-Lung-Risk model was developed using 147,497 CXRs of 40,643 asymptomatic smokers and never-smokers from the Prostate, Lung, Colorectal, and Ovarian (PLCO) cancer screening trial to predict lung-related mortality risk based on a single CXR image as input. In this study, we externally validated the model in a separate cohort of never-smokers having routine outpatient CXR from 2013-2014. The primary outcome was 6-year incident lung cancer, identified using International Classification of Disease (ICD) codes. Continuous CXR-Lung-Risk scores were converted to low, moderate, and high-risk groups based on externally derived risk thresholds.

RESULTS

Of 24,333 patients (mean age 63.4 ± 8.21 years; 44.3% male; 18,880 (80.5%) White, 1,789 (7.6%) Black, 789 (3.7%) Hispanic) included in the study, 32% (7774/24,333) were deemed high risk by CXR-Lung-Risk. 2.5% of the total cohort (616/24,333) developed lung cancer over 6 years of follow-up. CXR-Lung-Risk groups had a graded association with lung cancer risk, with 1.4% (37/2663) in the low-risk group (CXR-Lung-Risk <45), 2.2% (306/13896) in the moderate-risk group ($45 < \text{CXR-Lung-Risk} < 55$), and 3.5% (273/7774) in the high-risk group (CXR-Lung-Risk > 55). After adjusting for age, sex, race, previous lower respiratory tract infection, and prevalent COPD, there was still a 2.1 (95% CI [1.4,3.1]; $p < 0.001$) times greater risk of developing lung cancer in the high-risk group compared to low risk.

CONCLUSIONS

Using routine CXRs from the EMR, CXR-Lung-Risk identified never-smokers at high risk of lung cancer, a group in which lung cancer rates are increasing.

CLINICAL RELEVANCE/APPLICATIONS

CXR-Lung-Risk identified never-smokers at high risk of lung cancer, well above the >1.3% 6-year risk threshold where lung cancer screening CT is recommended by National Comprehensive Cancer Network guidelines.