








Development and Validation of a Risk Prediction Model for Second Primary Lung Cancer

Eunji Choi , PhD,¹ Nilotpall Sanyal, PhD,¹ Victoria Y. Ding, MS,¹ Rebecca M. Gardner , MS,¹ Jacqueline V. Aredo , BS,² Justin Lee, MPH,¹ Julie T. Wu, MD,² Thomas P. Hickey, BS,³ Brian Barrett, BS,³ Thomas L. Riley, BS,³ Lynne R. Wilkens, DrPH,⁴ Ann N. Leung, MD,⁵ Loïc Le Marchand , MD,⁴ Martin C. Tammemägi, PhD,⁶ Rayjean J. Hung , PhD,⁷ Christopher I. Amos , PhD,⁸ Neal D. Freedman, PhD,⁹ Iona Cheng, PhD,¹⁰ Heather A. Wakelee , MD,^{2,11} Summer S. Han, PhD^{1,11,12,*}

¹Quantitative Sciences Unit, Stanford University School of Medicine, Stanford, CA, USA; ²Stanford University School of Medicine, Stanford, CA, USA; ³Information Management Services, Rockville, MD, USA; ⁴Cancer Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA; ⁵Department of Radiology, Stanford University School of Medicine, Stanford, CA, USA; ⁶Department of Health Sciences, Brock University, St Catharines, Ontario, Canada; ⁷Prosserman Centre for Population Health Research, Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, Ontario, Canada; ⁸Department of Medicine, Baylor College of Medicine, Houston, TX, USA; ⁹Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA; ¹⁰Department of Epidemiology and Biostatistics, University of California, San Francisco, CA, USA; ¹¹Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA; and ¹²Department of Neurosurgery, Stanford University School of Medicine, Stanford, CA, USA

*Correspondence to: Summer S. Han, PhD, Departments of Medicine and of Neurosurgery, Stanford University School of Medicine, 1701 Page Mill Rd, Rm 234, Stanford, CA 94304, USA (e-mail: summer.han@stanford.edu).

Abstract

Background: With advancing therapeutics, lung cancer (LC) survivors are rapidly increasing in number. Although mounting evidence suggests LC survivors have high risk of second primary lung cancer (SPLC), there is no validated prediction model available for clinical use to identify high-risk LC survivors for SPLC. **Methods:** Using data from 6325 ever-smokers in the Multiethnic Cohort (MEC) study diagnosed with initial primary lung cancer (IPLC) in 1993-2017, we developed a prediction model for 10-year SPLC risk after IPLC diagnosis using cause-specific Cox regression. We evaluated the model's clinical utility using decision curve analysis and externally validated it using 2 population-based data—Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO) and National Lung Screening Trial (NLST)—that included 2963 and 2844 IPLC (101 and 93 SPLC cases), respectively. **Results:** Over 14 063 person-years, 145 (2.3%) ever-smoking IPLC patients developed SPLC in MEC. Our prediction model demonstrated a high predictive accuracy (Brier score = 2.9, 95% confidence interval [CI] = 2.4 to 3.3) and discrimination (area under the receiver operating characteristics [AUC] = 81.9%, 95% CI = 78.2% to 85.5%) based on bootstrap validation in MEC. Stratification by the estimated risk quartiles showed that the observed SPLC incidence was statistically significantly higher in the 4th vs 1st quartile (9.5% vs 0.2%; $P < .001$). Decision curve analysis indicated that in a wide range of 10-year risk thresholds from 1% to 20%, the model yielded a larger net-benefit vs hypothetical all-screening or no-screening scenarios. External validation using PLCO and NLST showed an AUC of 78.8% (95% CI = 74.6% to 82.9%) and 72.7% (95% CI = 67.7% to 77.7%), respectively. **Conclusions:** We developed and validated a SPLC prediction model based on large population-based cohorts. The proposed prediction model can help identify high-risk LC patients for SPLC and can be incorporated into clinical decision making for SPLC surveillance and screening.

Lung cancer (LC) is a major cause of cancer mortality. Recent studies showed that low-dose computed tomography (LDCT) screening reduces LC mortality by 16%-24% (1-3). With the adoption of LDCT and advancing therapeutics, the number of LC survivors is estimated at 571 340 in 2019, expected to rapidly increase to more than 724 000 by 2030 (4). However, recent

evidence suggests that LC survivors are at 4-6 times higher risk of developing second primary lung cancer (SPLC) vs the risk of initial primary lung cancer (IPLC) in the general population (5).

The US Preventive Services Task Force (USPSTF) established the national lung screening guidelines for IPLC for high-risk general population (6,7); however, no evidence-based consensus

Received: February 10, 2021; Revised: May 4, 2021; Accepted: July 12, 2021

© The Author(s) 2021. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

screening guidelines exist for LC survivors (8-10). Establishing efficient screening strategies for LC survivors requires the evaluation of the risk factors and prediction models for SPLC. Although prior studies (11-16), including ours (17), examined the potential risk factors and prediction models for SPLC, the major limitation is their reliance on limited domains of predictors without fully incorporating potential key factors, such as smoking exposure, treatment, or medical history. Furthermore, few existing modeling studies for SPLC have been externally validated that are available for clinical use.

In this study, we aimed to develop a risk prediction model for SPLC, incorporating various exposures based on a prospective population-based cohort, the Multiethnic Cohort (MEC) study. We evaluated the clinical utility of the prediction model in the context of risk-based screening and externally validated the model using 2 large-scale, randomized screening trials.

Methods

Study Population and Assessment of Risk Factors

The MEC is a population-based study from 5 racial and ethnic populations in California and Hawaii. Epidemiologic data were collected (1993-1996) through a self-reported questionnaire at enrollment. Our study cohort included participants (n = 6325) with an ever-smoking history who were diagnosed with IPLC in 1993-2017, identified via linkage to Surveillance, Epidemiology, and End Results (SEER) registries. Given that the goal of this study was to predict SPLC risk using the patient information collected at the time of IPLC diagnosis, we updated smoking-related variables with 10-year follow-up data (2003-2008) prior to IPLC diagnosis, if available (n = 1693; see [Supplementary Table 1](#), available online). As a potential predictor for SPLC risk, we evaluated a variable for meeting the USPSTF 2013 criteria (ie, aged 55-80 years, smoked ≥ 30 pack-years and ≤ 15 years since cessation for former smokers) (6) at the time of IPLC diagnosis. Initial tumor characteristics and treatment data were collected at IPLC diagnosis via linkage to SEER registries.

Outcome

The primary outcome was defined as the time from IPLC diagnosis to SPLC, death, or censored at the end of follow-up, whichever occurred first. SPLC was defined by the Martini and Melamed criteria (18).

Statistical Analysis

Competing Risk Regression. We applied a set of competing-risk models to obtain unbiased estimates of SPLC risks among LC patients, a substantial proportion of whom tend to die before developing SPLC because of high comorbidity. We applied a cause-specific Cox regression to build a prediction model for SPLC risk at the time of IPLC diagnosis (19), and we applied Fine-Gray regression (FGR) (20). All reported P values from cause-specific Cox regression and FGR are based on the 2-sided Wald test.

Variable Selection and Model Performance. The selection of the candidate predictors (21) that entered into model development was based on prior studies (11-16) and the previous risk model for SPLC developed using SEER (17) (see details in the [Supplementary Methods](#), available online). For evaluating model performance, we examined calibration, discrimination

using the time-varying area under the receiver operating characteristics (AUC) (22), and prediction accuracy using the Brier score that takes into account competing risks (23). The model evaluation was performed through an internal validation technique using 200 bootstrapped resamples to obtain bias-corrected estimation for predictive performance (24) in MEC. As several guidelines are targeted at patients diagnosed with early-stage IPLC for SPLC screening after intensive surveillance for recurrence (8-10), we evaluated the performance of the proposed model among early-stage IPLC cases as a sensitivity analysis.

Risk Stratification and Decision Curve Analysis. We considered an alternative way to evaluate a risk model—the ability to stratify a population into groups with distinct risks that can substantially affect the risk-benefit balance of screening (25). We divided the study population into 4 groups based on the estimated 10-year risk from the proposed model and then compared the observed cumulative incidence of SPLC for each quartile group using the 2-sided Gray method (26).

To evaluate the clinical utility of the prediction model, we applied decision curve analysis (DCA) by calculating the net-benefit of the model using the true- and false-positive rates under varied risk thresholds for screening (27). See details in the [Supplementary Methods](#) (available online).

Handling Missing Data

The missing rate of the variables included in our data was relatively low, mostly between 0% and 3% ([Supplementary Table 2](#), available online). Therefore, as primary analysis, we conducted a complete-case analysis (n = 5354; [Supplementary Table 3](#), available online) using the participants with complete data for the variables used for each model-building process. For sensitivity analysis, we performed multiple imputation (28) and applied Rubin rules to obtain the pooled estimates to calculate a single predicted risk score for each participant ([Supplementary Methods](#), available online).

Sensitivity Analysis

The model was evaluated across different subgroups defined by age at IPLC diagnosis, smoking status, or smoking pack-years to evaluate the robustness of model performance. To address the temporal gap between the smoking assessment and IPLC diagnosis in MEC, we projected smoking variables up to the time at IPLC diagnosis ([Supplementary Methods](#), available online). We first evaluated the proposed model by applying it to the projected smoking data (vs observed data at baseline or follow-up). Second, we reestimated the model parameters based on the projected smoking data to compare its model performance with the proposed model based on the observed smoking data.

Given that the current guidelines for LC screening are targeted at those who ever-smoked, and our aim was to help extend the guidelines to LC patients with a high risk of SPLC, we focused on ever-smoking IPLC patients in model development in MEC. However, we were also interested in evaluating the performance of the proposed model using the entire IPLC cases that include both ever-smoking and never-smoking patients in MEC as a sensitivity analysis. Last, we further explored the potential of an alternative model developed using data that excluded 2098 patients who died or were lost to follow-up within 6 months after IPLC diagnosis.

External Validation

We conducted external validation of the proposed model for SPLC using 2 large population-based screening trial datasets from the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO) and the National Lung Screening Trial (NLST). Among 2963 and 2844 ever-smoking IPLC cases in PLCO and NLST, 101 (3.4%) and 93 (3.3%) developed SPLC, respectively. Selections of the study cohorts are shown in the [Supplementary Methods](#) (available online). Model discrimination, calibration, and predictive accuracy were calculated by applying the proposed model to each external dataset.

Implementation of a Web-Based Tool for Risk Calculation

We implemented the proposed model into a user-friendly, web-based tool, called the Second Primary Lung Cancer Risk Assessment Tool (SPLC-RAT), which can assess an individual-level 5-, 10-, and 15-year risk of SPLC from the time of IPLC diagnosis. The model is provided as an open access application for free public use and is hosted at <https://splc.shinyapps.io/SPLC-RAT/>.

Results

Study Population

Among 6325 ever-smoking IPLC cases in MEC, 145 (2.3%) developed SPLC over 14 063 person- (interquartile range = 0.2-2.6) years ([Table 1](#); [Supplementary Figure 1, A](#), available online). The majority of SPLC cases (60.7%) was diagnosed within 5 years from IPLC diagnosis, with 12.4% of the cases diagnosed after 10 years from initial diagnosis ([Supplementary Figure 1, B](#), available online). The cohort comprised 25.2% Whites, 27.4% African Americans, 21.5% Japanese Americans, 13.0% Latinos, and 8.4% Native Hawaiians ([Table 1](#)).

SPLC Risk Prediction Model and Model Performance: MEC Development Cohort

The proposed SPLC prediction model included the following risk factors: IPLC histology, prior history of cancer, meeting the 2013 USPSTF eligibility criteria at IPLC diagnosis (ie, the USPSTF eligibility), smoking intensity (cigarettes per day), IPLC surgery, and IPLC stage ([Table 2](#)). The participants with complete data for these factors ($n = 5354$) were used for model building ([Supplementary Table 3](#), available online). One of the key risk factors for SPLC was the prior history of cancer (hazard ratio [HR] = 1.44, 95% confidence interval [CI] = 1.00 to 2.06; $P = .047$); patients with a prior history of cancer had a 44% increased SPLC risk. IPLC patients who met the USPSTF eligibility (HR = 1.74, 95% CI = 1.15 to 2.63; $P = .008$), underwent surgery (HR = 2.10, 95% CI = 1.23 to 3.59; $P = .007$), and were diagnosed with a large cell IPLC (vs a squamous cell IPLC) (HR = 2.01, 95% CI = 0.88 to 4.57; $P = .10$) had an increased SPLC risk. The effect of meeting the USPSTF eligibility was modified and reduced among advance-stage IPLC patients because of interaction.

Overall, the proposed model yielded a good calibration across different risk decile groups with high discriminative performance (AUC = 81.9%, 95% CI = 78.2% to 85.5%) and prediction accuracy (Brier score = 2.9, 95% CI = 2.4 to 3.3) ([Figure 1, A](#)) based on 200 bootstrapped resamples. The application of the proposed

model to the subgroup of early-stage IPLC patients showed an AUC of 70.1% (95% CI = 63.9% to 76.4%) ([Supplementary Figure 2](#), available online).

Risk Stratification and DCA

The result in [Figure 2, A](#) shows that the cumulative risk of SPLC varies by different percentiles of the estimated risks using the proposed model. When the study population was stratified by the quartiles of the estimated 10-year risk using the prediction model ([Figure 2, B](#)), the observed SPLC incidence was statistically significantly higher in the 4th vs 1st quartile (9.5% vs 0.2%; $P < .001$), demonstrating that the proposed model can potentially be useful in identifying high-risk patients for SPLC. The cumulative risk curves for SPLC, stratified by the key factors in the proposed model, are shown in [Supplementary Figure 3](#) (available online), demonstrating that patients who had a prior history of cancer, met the USPSTF eligibility, and were diagnosed with large cell or adenocarcinoma IPLC tend to have an increased SPLC risk over time.

DCA ([Figure 2C](#)) identified a range of risk thresholds that are clinically useful in the context of screening; the results show that in a wide range of 10-year risk thresholds from 1% to 20%, the prediction model yields a larger net-benefit vs hypothetical all-screening or no-screening scenarios.

Sensitivity Analysis

Overall, we observed a good performance of the model across different subgroups in MEC ([Supplementary Figure 4](#), available online). Sensitivity analyses using FGR for estimating model parameters ([Supplementary Table 4](#) and [Supplementary Figure 5](#), available online) and based on multiple imputation ([Supplementary Table 5](#) and [Supplementary Figure 6](#), available online) showed consistent results, confirming internal validity based on bootstrapping. The model performance using the projected smoking data at IPLC diagnosis showed robust results ([Supplementary Table 6](#) and [Supplementary Figure 7](#), available online).

Besides the ever-smoking cohort used for model development, we identified 740 never-smoking IPLC patients in MEC ([Supplementary Table 7](#), available online). The performance of the proposed model using ever-smoking and never-smoking patients showed an AUC of 80.4% (95% CI = 76.3% to 84.8%) ([Supplementary Figure 8](#), available online).

Exploratory analyses were performed by excluding 2098 IPLC patients with less than 6 months follow-up after IPLC diagnosis. Despite losing nearly one-third of the data, the reestimated model ([Supplementary Table 8](#), available online) offered promising results with regards to discrimination and calibration ([Supplementary Figure 9](#), available online).

External Validation Using PLCO and NLST

Patient characteristics of the 2 external validation cohorts from PLCO and NLST are given in [Table 3](#) and [Supplementary Tables 9-10](#) (available online). Compared with MEC, the mean pack-years was higher in NLST and PLCO (63.6 and 57.8 for NLST and PLCO, respectively, vs 31.0 in MEC). External validation of the proposed model using PLCO and NLST showed good calibration and discrimination with an AUC of 78.8% (95% CI = 74.6% to 82.9%) and 72.7% (95% CI = 67.7% to 77.7%), respectively ([Figure 1](#)).

Table 1. Characteristics of study population stratified by outcome status in the Multiethnic Cohort

Variables	Total	Outcome			
		SPLC	IPLC death	Other death	Censored
Total events, No. (%)	6325 (100.0)	145 (2.3)	4093 (64.7)	1049 (16.6)	1038 (16.4)
Follow-up time, y					
Mean (IQR)	2.2 (0.2-2.6)	4.6 (1.0-6.8)	1.0 (0.2-1.2)	2.2 (0.2-3.0)	6.7 (3.0-9.3)
Demographic information					
Age at IPLC diagnosis					
Mean (SD), y	74.2 (8.2)	72.1(8.2)	73.7 (8.1)	74.1 (8.0)	76.2 (8.6)
Age groups, No. (%), y					
<55 y	113 (1.8)	3 (2.1)	78 (1.9)	20 (1.9)	12 (1.2)
55-59 y	234 (3.7)	9 (6.2)	170 (4.2)	33 (3.1)	22 (2.1)
60-69 y	1477 (23.4)	44 (30.3)	976 (23.8)	249 (23.7)	208 (20.0)
70-79 y	2950 (46.6)	68 (46.9)	1957 (47.8)	494 (47.1)	431 (41.5)
≥80 y	1551 (24.5)	21 (14.5)	912 (22.3)	253 (24.1)	365 (35.2)
Sex, No. (%)					
Female	2529 (40.0)	66 (45.5)	1564 (38.2)	401 (38.2)	498 (48.0)
Male	3796 (60.0)	79 (54.5)	2529 (61.8)	648 (61.8)	540 (52.0)
Race, No. (%)					
White	1591 (25.2)	43 (29.7)	1049 (25.6)	254 (24.2)	245 (23.6)
Japanese American	1357 (21.5)	34 (23.4)	880 (21.5)	223 (21.3)	220 (21.2)
African American	1736 (27.4)	38 (26.2)	1121 (27.4)	306 (29.2)	271 (26.1)
Latino	824 (13.0)	15 (10.3)	497 (12.1)	121 (11.5)	191 (18.4)
Native Hawaiian	533 (8.4)	11 (7.6)	363 (8.9)	90 (8.6)	69 (6.6)
Others	284 (4.5)	4 (2.8)	183 (4.5)	55 (5.2)	42 (4.0)
Education, No. (%)					
High school	3258 (51.5)	62 (42.8)	2162 (52.8)	561 (53.5)	473 (45.6)
College	2522 (39.9)	68 (46.9)	1589 (38.8)	403 (38.4)	462 (44.5)
Postgraduate	512 (8.1)	15 (10.3)	321 (7.8)	78 (7.4)	98 (9.4)
Data missing	61 (1.0)	0 (0.0)	21 (0.5)	7 (0.7)	33 (3.2)
Tumor characteristics					
Stage of IPLC, No. (%) ^a					
Early stage	2525 (39.9)	134 (92.4)	1214 (29.7)	532 (50.7)	645 (62.1)
Advanced stage	3414 (54.0)	10 (6.9)	2599 (63.5)	457 (43.6)	348 (33.5)
Data missing	386 (6.1)	1 (0.7)	280 (6.8)	60 (5.7)	45 (4.3)
Histology of IPLC, No. (%)					
Squamous cell	1369 (21.6)	35 (24.1)	842 (20.6)	252 (24.0)	240 (23.1)
Adenocarcinoma	2348 (37.1)	81 (55.9)	1364 (33.3)	390 (37.2)	513 (49.4)
Large cell	194 (3.1)	9 (6.2)	124 (3.0)	38 (3.6)	23 (2.2)
Small cell	720 (11.4)	3 (2.1)	554 (13.5)	87 (8.3)	76 (7.3)
Non-small cell carcinoma, NOS	504 (8.0)	4 (2.8)	390 (9.5)	73 (7.0)	37 (3.6)
Others ^b	1190 (18.8)	13 (9.0)	819 (20.0)	209 (19.9)	149 (14.4)
Smoking-related factors ^c					
Smoking status, No. (%)					
Former	3352 (53.0)	73 (50.3)	2119 (51.8)	570 (54.3)	590 (56.8)
Current	2973 (47.0)	72 (49.7)	1974 (48.2)	479 (45.7)	448 (43.2)
Smoking intensity					
Mean (SD), cigarettes per day	17.7 (8.7)	19.8 (9.4)	17.9 (8.7)	17.7 (8.6)	16.5 (8.8)
Data missing, No. (%)	137 (2.2%)	1 (0.7)	86 (2.1)	27 (2.6)	23 (2.2)
Pack-years					
Mean (SD)	31.0 (18.4)	35.3 (19.3)	31.8 (18.5)	30.3 (18.0)	27.8 (18.0)
Data missing, No. (%)	200 (3.2)	2 (1.4)	125 (3.1)	37 (3.5)	36 (3.5)
Met the 2013 USPSTF criteria, No. (%) ^d					
No	4077 (64.5)	72 (49.7)	2546 (62.2)	685 (65.3)	774 (74.6)
Yes	2137 (33.8)	71 (49.0)	1475 (36.0)	341 (32.5)	250 (24.1)
Data missing	111 (1.8)	2 (1.4)	72 (1.8)	23 (2.2)	14 (1.3)
Quit-years ^e					
Mean (SD)	10.8 (6.0)	9.6 (6.9)	10.5 (6.0)	10.9 (6.9)	12.0 (6.5)
Data missing, No. (%)	48 (0.8)	0 (0.0)	26 (0.6)	11 (1.0)	11 (1.1)
Clinical factors					
Prior history of cancer, No. (%) ^f					
No	4712 (74.5)	98 (67.6)	3142 (76.8)	717 (68.4)	755 (72.7)
Yes	1613 (25.5)	47 (32.4)	951 (23.2)	332 (31.6)	283 (27.3)

(continued)

Table 1. (continued)

Variables	Total	Outcome			
		SPLC	IPLC death	Other death	Censored
BMI					
Mean (SD), kg/m ²	25.9 (4.7)	26.4 (4.7)	25.8 (4.5)	26.2 (5.2)	26.3 (4.7)
Data missing, No. (%)	76 (1.2)	0 (0.0)	50 (1.2)	16 (1.5)	10 (1.0)
Treatments for IPLC					
Radiotherapy, No. (%)					
No	3976 (62.9)	122 (84.1)	2356 (57.6)	752 (71.7)	746 (71.9)
Yes	2180 (34.5)	23 (15.9)	1616 (39.5)	277 (26.4)	264 (25.4)
Data missing	169 (2.7)	0 (0.0)	121 (3.0)	20 (1.9)	28 (2.7)
Chemotherapy, No. (%)					
No	3964 (62.7)	129 (89.0)	2403 (58.7)	748 (71.3)	684 (65.9)
Yes	2059 (32.6)	15 (10.3)	1483 (36.2)	244 (23.3)	317 (30.5)
Data missing	302 (4.8)	1 (0.7)	207 (5.1)	57 (5.4)	37 (3.6)
Surgery, No. (%)					
No	4458 (70.5)	25 (17.2)	3226 (78.8)	657 (62.6)	550 (53.0)
Yes	1313 (20.8)	112 (77.2)	465 (11.4)	282 (26.9)	454 (43.7)
Data missing	554 (8.8)	8 (5.5)	402 (9.8)	110 (10.5)	34 (3.3)

^aDisease extent was defined using Surveillance, Epidemiology, and End Results Extent of Disease as local and regional for early stage and distant for advanced stage. BMI = body mass index; IPLC = initial primary lung cancer; IQR = interquartile range; NOS = not otherwise specified; SPLC = second primary lung cancer; USPSTF = US Preventive Services Task Force.

^bClassification of "other" histology based on International Classification of Diseases–O-3 codes including 8000, 8001, 8010, 8020, 8022, 8030-8033, 8200, 8240, 8244, 8246, 8249, 8560, 8720, 8800, 8810, and 8980; all confirmed lung cancer diagnosis.

^cSmoking data were updated with available 10-year follow-up information close or prior to IPLC diagnosis for 26.8%.

^dAged 55-80 years, smoked ≥ 30 pack-years of smoking, and ≤ 15 years since cessation.

^eAmong former smokers only.

^fHistory of cancer, other than lung, before the time of IPLC diagnosis.

Using the 75th percentile risk threshold derived from MEC (ie, the 4th quartile with an estimated 10-year risk larger than 2.8%), 37% and 48% of the patients in PLCO and NLST were in a high-risk subgroup, in which the observed 10-year SPLC incidence was 7.7% and 9.6% for PLCO and NLST, respectively (data not shown).

We implemented the proposed model for predicting SPLC risk that was externally validated into SPLC-RAT (Supplementary Figure 10, available online).

Discussion

We developed a prediction model for SPLC risk among ever-smoking IPLC cases that integrates various risk factors for SPLC, using large population-based cohort data. The proposed model was validated using 2 external populations that are heterogeneous with regards to smoking history and race and ethnicity. The development MEC cohort included relatively light smokers (mean pack-years = 31.0) vs the validation cohorts, PLCO (57.8 pack-years), and NLST (63.6 pack-years); in addition, MEC is diverse with a high percentage of Asians and African Americans, but the validation cohorts are overwhelmingly White. However, our model, built using MEC, performed well in both cohorts, yielding an AUC of more than 70% and thus demonstrating the potential generalizability of the proposed model in predicting SPLC risk.

The proposed model includes several predictors that were previously unexamined (13-17), including the prior history of cancer, the USPSTF eligibility, and smoking intensity. Other smoking variables such as pack-years, duration, and quit-years were not directly included in the model but captured by the USPSTF eligibility, a composite measure of smoking and age criteria, which was shown to be a stronger predictor for SPLC

based on our analysis. Thus, our results suggest that smoking still plays an important role in predicting SPLC risk among IPLC patients. The prior history of cancer was associated with an increased SPLC risk in our analysis, which was consistent with the results for other second malignancies (29). Our exploratory analysis focusing on the prior history of smoking-related cancer showed a stronger association with SPLC; however, the difference in model performance was minimal vs any type of prior cancer (data not shown). Race and ethnicity is one of the main predictors for IPLC risk, with African Americans showing increased risks of LC incidence and mortality (30,31). However, our analysis did not find statistically significant racial and ethnic effects on SPLC risk. One potential explanation is that once diagnosed with IPLC, the risk of subsequent SPLC is no longer associated with race and ethnicity.

To the best of our knowledge, the proposed model for SPLC presents the first effort to incorporate comprehensive risk factors including smoking information, medical history, treatment, and tumor characteristics in predicting SPLC risk using large, population-based data. The proposed model was validated by using 2 external population-based datasets with an AUC ranging from 72.7% to 78.8%. Thorough sensitivity analyses were performed to examine the predictive accuracy of the proposed model, which largely showed consistent findings. We implemented the proposed model into a user-friendly, web-based tool—SPLC-RAT—that can help decision making of patients and clinicians for guiding surveillance and screening for IPLC cases.

Regarding the utility of SPLC-RAT in clinical practice, an oncologist could use this assessment tool after IPLC diagnosis or treatment by applying a certain risk threshold. For example, if the estimated risk of a patient is higher than a risk threshold, for instance, a 10-year SPLC risk of 2.8% (derived from the 75th percentile risk threshold in the development cohort), the patient can be recommended to receive LDCT screening frequently

Table 2. Cause-specific hazard ratios of risk factors for second primary lung cancer in the final proposed risk prediction model in the Multiethnic Cohort^a

Factors	No.	Cause-specific Cox hazards model	
		HR (95% CI)	p ^b
Histology of IPLC			
Squamous cell	1185	Referent	
Large cell	2053	2.01 (0.88 to 4.57)	.01
Adenocarcinoma	163	1.15 (0.76 to 1.75)	.51
Small cell	624	0.79 (0.23 to 2.66)	.70
Non-small cell carcinoma, NOS	473	0.88 (0.30 to 2.57)	.82
Other ^c	856	0.99 (0.52 to 1.89)	.97
Prior history of cancer^d			
No	3949	Referent	
Yes	1405	1.44 (1.00 to 2.06)	.047
Met the 2013 USPSTF criteria^e			
No	3539	Referent	
Yes	1815	1.74 (1.15 to 2.63)	.008
Smoking intensity, cigarettes per day	5354	1.01 (0.99 to 1.04)	.25
Surgery for IPLC			
No	2525	Ref	
Yes	3414	2.10 (1.23 to 3.59)	.007
Stage of IPLC			
Early stage ^f	2254	Referent	
Advanced stage	3100	0.48 (0.21 to 1.07)	.07
Stage of IPLC × Met the 2013 USPSTF Criteria ^e		0.28 (0.06 to 1.36)	.11

^aBased-on participants (n = 5354) with complete data for the variables in the proposed prediction model. CI = confidence interval; HR = hazard ratio; IPLC = initial primary lung cancer; NOS = not otherwise specified; USPSTF = US Preventive Services Task Force.

^bp value by 2-sided Wald for the cause-specific Cox hazards model.

^cClassification of "other" histology based on International Classification of Diseases-O-3 codes including 8000, 8001, 8010, 8020, 8022, 8030-8033, 8200, 8240, 8244, 8246, 8249, 8560, 8800, 8810, 8890, and 8980; all confirmed lung cancer diagnosis.

^dHistory of cancer, other than lung, before the time of IPLC diagnosis.

^eAged 55-80 years, smoked ≥30 pack-years of smoking, and ≤15 years since cessation.

^fDisease extent was defined using Surveillance, Epidemiology, and End Results Extent of Disease as local and regional for early stage and distant for advanced stage.

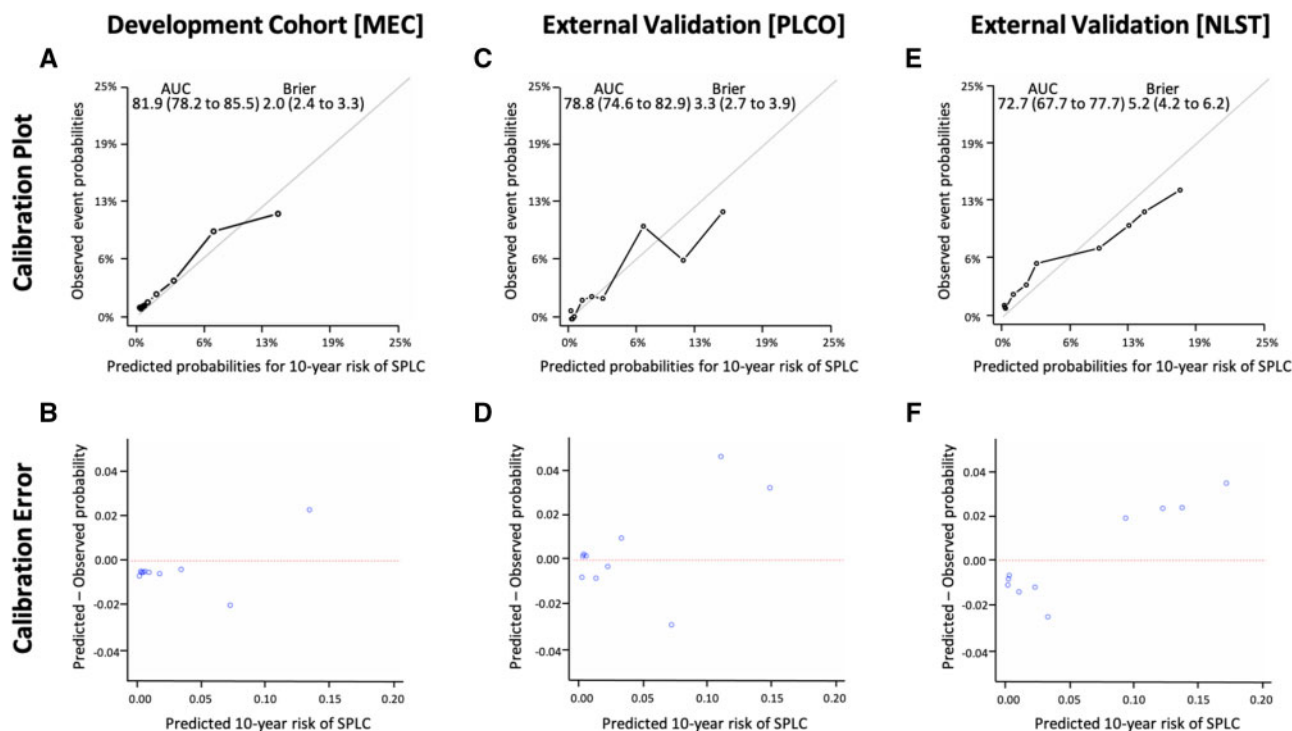


Figure 1. Performance of the proposed risk prediction model for second primary lung cancer. **A)** Calibration plots with discriminative performance (area under the curve [AUC]) and prediction accuracy (Brier score) based on an internal validation using 200 bootstrapped resamples from the development cohort (the Multiethnic Cohort [MEC]). **B)** Plots of mean difference from predicted to observed probability (calibration error) across risk deciles from the MEC. Calibration plots and calibration errors from 2 external validation datasets of the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO) in (C) and (D) and of the National Lung Screening Trials (NLST) in (E) and (F), respectively. 95% confidence intervals are shown within parentheses for the AUCs and Brier scores.

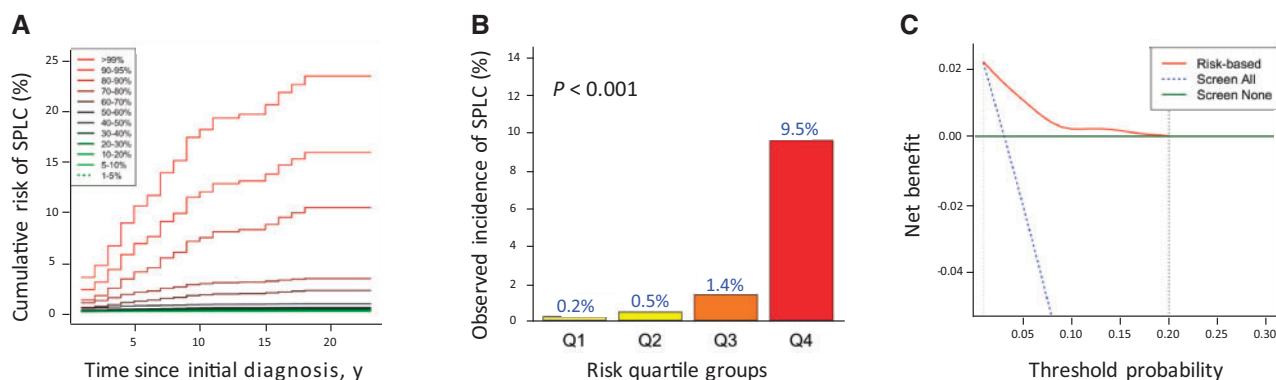


Figure 2. Risk stratification for second primary lung cancer and decision curve analysis (DCA) in the Multiethnic Cohort (MEC). **A**) Cumulative risk of second primary lung cancer (SPLC) by percentiles of the estimated risk on the basis of the proposed risk prediction model. **B**) Cumulative incidence of SPLC by quartiles of the estimated 10-year risk of SPLC using the proposed risk-prediction model. The 4 quartile groups were divided based on the following risk thresholds: Q1: $r \leq 0.3\%$; Q2: $0.3\% < r \leq 0.5\%$; Q3: $0.5\% < r \leq 2.8\%$; and Q4: $r > 2.8\%$. The risks across different groups were compared and tested using the method by 2-sided Gray test. **C**) Decision curve analysis for the proposed risk prediction model. In DCA, the red curve depicts the net benefit of the risk model-based selection strategy for screening across different 10-year risk thresholds (from 1% to 20% shown in the x-axis), whereas the blue and green lines display the net benefits in the alternative strategies of screening all or none. The analyses in A, B, and C were based on 5354 participants with complete data for 6 variables in the proposed risk prediction model (stage and histology of initial primary lung cancer, surgery for initial primary lung cancer, prior history of cancer, smoking intensity measured by cigarettes per day, and the US Preventive Services Task Force eligibility).

Table 3. Characteristics of external study population from the PLCO and the NLST^a

Variables	PLCO		NLST	
	SPLC No. (%)	All No. (%)	SPLC No. (%)	All No. (%)
Total No. of events (%)	101 (100.0)	2963 (100.0)	93 (100.0)	2844 (100.0)
Follow-up time, y				
Mean (SD)	5.5 (3.8)	2.5 (3.5)	4.5 (2.7)	2.6 (3.0)
Demographic information				
Age at IPLC diagnosis				
Mean (SD), y	69.1 (6.1)	71.5 (6.4)	65.5 (5.9)	68.4 (6.1)
Age groups, No. (%)				
55-59 y	7 (6.9)	99 (3.3)	17 (18.3)	202 (7.1)
60-69 y	49 (48.5)	1036 (35.0)	51 (54.8)	1433 (50.4)
70-79 y	40 (39.6)	1508 (50.9)	24 (25.8)	1102 (38.7)
≥80 y	5 (5.0)	320 (10.8)	1 (1.1)	107 (3.8)
Sex, No. (%)				
Female	39 (38.6)	1132 (38.2)	43 (46.2)	1202 (42.3)
Male	62 (61.4)	1831 (61.8)	50 (53.8)	1642 (57.7)
Race, No. (%)				
White	99 (97.0)	2620 (88.4)	78 (83.9)	2578 (90.6)
Asian	0 (0.0)	70 (2.4)	4 (4.3)	45 (1.6)
African American	2 (2.0)	207 (7.0)	9 (9.7)	141 (5.0)
Latino	0 (0.0)	38 (1.3)	1 (1.1)	34 (1.2)
Others ^b	1 (1.0)	28 (1.0)	1 (1.1)	46 (1.6)
Education, No. (%)				
High school	60 (59.4)	1565 (52.8)	43 (46.2)	1510 (53.1)
College	32 (31.7)	1109 (37.4)	37 (39.8)	1007 (35.4)
Postgraduate	9 (8.9)	285 (9.6)	11 (11.8)	276 (9.7)
Data missing	0 (0.0)	4 (0.1)	2 (2.2)	51 (1.8)
Tumor characteristics				
Stage of IPLC, No. (%) ^c				
Early stage	92 (91.1)	1625 (54.8)	88 (94.6)	1761 (61.6)
Advanced stage	9 (8.9)	1338 (45.2)	5 (5.4)	1093 (38.4)
Histology of IPLC, No. (%)				
Squamous cell	28 (27.7)	668 (22.5)	25 (26.9)	650 (22.9)
Adenocarcinoma	53 (52.5)	1256 (42.4)	44 (47.3)	1173 (41.2)
Large cell	4 (4.0)	117 (3.9)	3 (3.2)	95 (3.3)
Small cell	5 (5.0)	446 (15.1)	6 (6.5)	455 (16.0)
Non-small cell carcinoma, NOS, No. (%)	0 (0.0)	0 (0.0)	10 (10.8)	289 (10.2)
Others ^d	11 (10.8)	476 (16.1)	5 (5.4)	182 (6.4)

(continued)

Table 3. (continued)

Variables	PLCO		NLST	
	SPLC No. (%)	All No. (%)	SPLC No. (%)	All No. (%)
Smoking-related factors				
Smoking status, No. (%)				
Former	53 (52.5)	1744 (58.9)	35 (37.6)	1098 (38.6)
Current	48 (47.5)	1219 (41.1)	58 (62.4)	1746 (61.4)
Smoking intensity ^a				
Mean (SD), cigarettes/d	26.4 (13.7)	24.9 (12.1)	30.4 (10.5)	29.2 (11.5)
Pack-years ^e				
Mean (SD)	62.8 (34.4)	57.8 (32.8)	64.7 (22.3)	63.5 (26.2)
Met the 2013 USPSTF criteria, No. (%) ^f				
No	30 (29.7)	1266 (42.7)	3 (3.2)	332 (11.7)
Yes	71 (70.3)	1697 (57.3)	90 (96.8)	2512 (88.3)
Quit-years				
Mean (SD)	8.8 (12.2)	11.4 (13.4)	3.5 (5.8)	4.4 (6.7)
Clinical factors				
Prior history of cancer, No. (%) ^g				
No	94 (93.1)	2735 (92.3)	87 (93.5)	2671 (93.9)
Yes	7 (6.9)	228 (7.7)	6 (6.5)	173 (6.1)
BMI				
Mean (SD), kg/m ²	26.3 (3.6)	26.6 (4.5)	27.8 (6.1)	26.9 (4.8)
Data missing, No. (%)	1 (1.0)	38 (1.3)	1 (1.1)	14 (0.5)
Treatments for IPLC				
Radiotherapy, No. (%)				
No	77 (76.2)	1754 (59.2)	72 (77.4)	1792 (63.0)
Yes	24 (23.8)	1209 (40.8)	21 (22.6)	1037 (36.5)
Data missing	0 (0.0)	0 (0.0)	0 (0.0)	15 (0.5)
Chemotherapy, No. (%)				
No	73 (72.3)	1444 (48.7)	60 (64.5)	1414 (49.7)
Yes	28 (27.7)	1519 (51.3)	33 (35.5)	1406 (49.4)
Data missing	0 (0.0)	0 (0.0)	0 (0.0)	24 (0.8)
Surgery, No. (%)				
No	17 (16.8)	2018 (68.1)	15 (16.1)	1632 (57.4)
Yes	84 (83.2)	945 (31.9)	78 (83.9)	1212 (42.6)

^aThe tables stratified by full outcome status are shown in [Supplementary Tables 9 and 10](#) (available online). BMI = body mass index; IPLC = initial primary lung cancer; NLST = National Lung Screening Trial; PLCO = Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; SPLC = second primary lung cancer; USPSTF = US Preventive Services Task Force.

^bOthers included races from Pacific Islander, American Indian, and missing values.

^cDisease extent was defined using Surveillance, Epidemiology, and End Results Extent of Disease as local and regional for early stage and distant for advanced stage.

^dClassification of "other" histology based on International Classification of Diseases–O-3 codes including 8000, 8010, 8012, 8020, 8022, 8031, 8032, 8033, 8246, and 8560 in the PLCO and 8000, 8001, 8010, 8022, 8032, 8033, 8230, 8240, 8249, 8560, and 8980 in the NLST; all confirmed lung cancer cases.

^eSmoking data from the PLCO closest to initial diagnosis were extracted.

^fAged 55-80 years, smoked ≥ 30 pack-years of smoking, and ≤ 15 years since cessation.

^gHistory of cancer, other than lung, before the time of IPLC diagnosis.

(ie, every 6 months). Otherwise, less frequent screening, such as annual or biennial, could be suggested. However, selection of optimal risk thresholds and screening intervals for SPLC warrants further investigation, which potentially involves microsimulation-based decision analyses (32,33).

Our study is not without limitations. Unlike our previous SPLC model using SEER (17) that focused on long-term survivors (≥ 5 years since IPLC diagnosis), the present model did not exclude short-term survivors who died within 5 years since IPLC diagnosis because of the decreased number of SPLC (from 135 to 48 SPLC cases and from 5354 to 739 IPLC cases). Along the same lines, prediction models for SPLC could be more useful when targeting patients who are more likely to survive, such as surgically treated patients for IPLC, which would also lead to a large reduction in our sample size ($>70\%$). However, the application of the proposed model to the subgroup of early-stage IPLC patients—who can potentially survive longer than the

others—showed a moderate AUC (70.1%, 95% CI = 63.9% to 76.4%) and stronger risk stratification. We also explored an alternative model using data that excluded patients with less than 6 months follow-up after IPLC diagnosis, which can help examine the potential of predictive modeling for SPLC targeted for survivors of LC; the analysis demonstrated promising results with an AUC of 73.1% (95% CI = 66.4% to 78.3%) despite losing nearly one-third of the data. Our future directions include the development of a comprehensive model for SPLC using a sufficient number of long-term survivors. As with any cohort data, MEC may have selection bias, occurring from recruiting participants, collecting exposure information, and following up participants. Also, the IPLC cases in MEC have relatively lower smoking prevalence (40.7%) compared with previous studies with LC patients (42%-83%) (34), with disproportionately higher proportions of Asian and African Americans. However, the proposed model was validated using heterogeneous external

populations under varying smoking prevalence levels (41.3% for PLCO and 61.5% for NLST) and different race and ethnic distributions. Furthermore, MEC helped increase the power of the study as a primary cohort by providing the largest number of SPLC cases with long-term follow-up and detailed exposure data among all cohorts considered. We used a binary IPLC staging variable in predicting SPLC risk, merging patients with stage IA to IIIA IPLCs into an early-stage IPLC group based on the literature (17,35); also, analyses using SEER showed that SPLC risk does not vary statistically significantly across stage IA to IIIA IPLCs (data not shown). Therefore, the clinical utility of the proposed model would not be diminished given that the model continues to capture well the SPLC risk variations using parsimonious dichotomous staging parameters. The proposed SPLC model focused on ever-smokers, excluding never-smokers, given the lack of evidence on the efficacy of LDCT among never-smokers who are excluded from the current screening guidelines. Still, our sensitivity analysis showed a relatively good performance of the proposed model in the entire MEC cohort that included never-smokers. Although we used the well-established criteria for SPLC (18), this definition is restricted compared with those used in practical clinical settings, where SPLC cases are more frequently identified radiographically. Evaluation of the model performance using data that incorporate additional radiographic criteria for SPLC merits further investigation.

To conclude, we presented a prediction model for SPLC that demonstrated good discrimination, calibration, and predictive accuracy as confirmed across different population-based data. The web-based implementation for risk-assessment can provide a potentially useful tool for evaluating an individual's risk of SPLC and can help decision making of patients and reduce clinicians' uncertainty on how to best guide LC survivors for screening and surveillance for SPLC.

Funding

This study is supported by grant from the National Institutes of Health (1R37CA226081).

Notes

Role of the funder: The funder had no role in study design, data collection, data interpretation, or writing of the report.

Disclosures: Dr Tammemagi is a consultant for Johnson & Johnson/Janssen, Medial EarlySign, Nucleix, bioAffinity Technologies, and AstraZeneca. Dr Wakelee reports grants from Gilead; personal fees from Janssen, Daiichi Sankyo, INC, Helsinn, Mirati, UpToDate; personal fees and non-financial support from AztraZeneca; personal fees and research funding to the institution from Xcovery; non-financial support from Takeda, CellWorks, Clinical Care Options Oncology LLC, Fishawack Facilitate LTD, Medscape, Onclive/Intellisphere LLC, Phillips Gilmore Oncology, Physicians Education Resource LLC/MJH (Targeted Oncology), Potomac Center for Medical Education (Rockpointe), Prime Oncology LLC, Primo, Research to Practice, WebMD Health, Novartis, RGCON—Rajiv Gandhi Conference, J LCS—Japanese Lung Cancer Society, KSMO—Korean Society of Medical Oncology, Stanford University, ITMIG; non-financial support and research funding to the institution from Genentech/Roche, Merck; research funding to the institution from ACEA Biosciences, Arrys Therapeutics, AztraZeneca/

MedImmune, BMS, Celgene, Clovis Oncology, Exelixis, Lilly, Pfizer, Pharmacyclics, all outside the submitted work. All remaining authors report no other disclosures.

Author contributions: Conceptualization, SSH, HAW; Data Curation, SSH, EC, LRW, TPH, BB, TLR; Formal Analysis, SSH, EC, NS, VYD, RMG, JVA, JL, TPH, BB, TLR; Writing—original draft, EC; Writing—review & editing, all authors; Validation, all authors; Supervision, SSH.

Acknowledgements: We appreciate the support that Natasha Purington provided for creating the framework of R Shiny application development. Cancer incidence data have been provided by the Alabama Statewide Cancer Registry, Arizona Cancer Registry, Colorado Central Cancer Registry, District of Columbia Cancer Registry, Georgia Cancer Registry, Hawaii Cancer Registry, Cancer Data Registry of Idaho, Maryland Cancer Registry, Michigan Cancer Surveillance Program, Minnesota Cancer Surveillance System, Missouri Cancer Registry, Nevada Central Cancer Registry, Ohio Cancer Incidence Surveillance System, Pennsylvania Cancer Registry, Texas Cancer Registry, Utah Cancer Registry, Virginia Cancer Registry, and Wisconsin Cancer Reporting System. All are supported in part by funds from the Center for Disease Control and Prevention, National Program for Central Registries, local states, or by the National Cancer Institute Surveillance, Epidemiology, and End Results program. The results reported here and the conclusions derived are the sole responsibility of the authors.

Data Availability

The data underlying this analysis were provided by the Multiethnic Cohort Study (MEC) under data use agreement. Researchers interested in the MEC data may submit an inquiry online: <https://www.uhcancercenter.org/for-researchers/mec-data-sharing>.

References

- de Koning HJ, van der Aalst CM, de Jong PA, et al. Reduced lung-cancer mortality with volume CT screening in a randomized trial. *N Engl J Med*. 2020; 382(6):503–513.
- Aberle DR, Adams AM, Berg CD, et al.; for the National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011;365(5):395–409.
- Pinsky PF, Church TR, Izmirlian G, Kramer BS. The National Lung Screening Trial: results stratified by demographics, smoking history, and lung cancer histology. *Cancer*. 2013;119(22):3976–3983.
- Miller KD, Nogueira L, Mariotto AB, et al. Cancer treatment and survivorship statistics, 2019. *CA A Cancer J Clin*. 2019;69(5):363–385.
- Surapaneni R, Singh P, Rajagopalan K, Hageboutros A. Stage I lung cancer survivorship: risk of second malignancies and need for individualized care plan. *J Thorac Oncol*. 2012;7(8):1252–1256.
- Moyer VA; for the U.S. Preventive Services Task Force. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2014;160(5):330–338.
- Krist AH, Davidson KW, Mangione CM, et al. Screening for Lung Cancer: U.S. Preventive Services Task Force Recommendation Statement. *JAMA*. 2021;325(10):962–970.
- National Comprehensive Cancer Network. *Non-Small Cell Lung Cancer*. In: NCCN Clinical Practice Guidelines in Oncology version 8; 2020.
- Schneider BJ, Ismaila N, Aerts J, et al. Lung cancer surveillance after definitive curative-intent therapy: ASCO guideline. *J Clin Oncol*. 2020;38(7):753–766.
- Postmus PE, Kerr KM, Oudkerk M, et al.; for the ESMO Guidelines Committee. Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2017;28(suppl 4):iv1–iv21.
- Boyle JM, Tandberg DJ, Chino JP, et al. Smoking history predicts for increased risk of second primary lung cancer: a comprehensive analysis. *Cancer*. 2015; 121(4):598–604.

12. Ripley RT, McMillan RR, Sima CS, et al. Second primary lung cancers: smokers versus nonsmokers after resection of stage I lung adenocarcinoma. *Ann Thorac Surg*. 2014;98(3):968–974.
13. Lou F, Huang J, Sima CS, et al. Patterns of recurrence and second primary lung cancer in early-stage lung cancer survivors followed with routine computed tomography surveillance. *J Thorac Cardiovasc Surg*. 2013;145(1):75–81; discussion 81–72.
14. Thakur MK, Ruterbusch JJ, Schwartz AG, et al. Risk of second lung cancer in patients with previously treated lung cancer: analysis of Surveillance, Epidemiology, and End Results (SEER) data. *J Thorac Oncol*. 2018;13(1):46–53.
15. Reinmuth N, Stumpf A, Stumpf P, et al. Characteristics and outcome of patients with second primary lung cancer. *Eur Respir J*. 2013;42(6):1668–1676.
16. Spratt DE, Wu AJ, Adeseye V, et al. Recurrence patterns and second primary lung cancers after stereotactic body radiation therapy for early-stage non-small-cell lung cancer: implications for surveillance. *Clin Lung Cancer*. 2016;17(3):177–183.e172.
17. Han SS, Rivera GA, Tammemägi MC, et al. Risk stratification for second primary lung cancer. *J Clin Oncol*. 2017;35(25):2893–2899.
18. Martini N, Melamed MR. Multiple primary lung cancers. *J Thorac Cardiovasc Surg*. 1975;70(4):606–612.
19. Prentice RL, Kalbfleisch JD, Peterson AV Jr, et al. The analysis of failure times in the presence of competing risks. *Biometrics*. 1978;34(4):541–554.
20. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999;94(446):496–509.
21. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J*. 2014;35(29):1925–1931.
22. Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med*. 2013;32(30):5381–5397.
23. Gerds TA, Andersen PK, Kattan MW. Calibration plots for risk prediction models in the presence of competing risks. *Stat Med*. 2014;33(18):3191–3203.
24. Harrell FE. *Regression Modeling Strategies*. Switzerland: Springer; 2014.
25. Chatterjee N, Shi J, Garcia-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat Rev Genet*. 2016;17(7):392–406.
26. Gray RJ. A class of K-sample tests for comparing the cumulative incidence of a competing risk. *Ann Statist*. 1988;16(3):1141–1154.
27. Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Med Inform Decis Mak*. 2008;8:53.
28. Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*. 2009;338:b2393.
29. Henderson LM, Durham DD, Tammemägi MC, et al. Lung cancer screening with low dose computed tomography in patients with and without prior history of cancer in the National Lung Screening Trial. *J Thorac Oncol*. 2021;16(6):980–989.
30. Tammemägi MC, Church TR, Hocking WG, 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.
31. Tammemägi MC, Katki HA, Hocking WG, et al. Selection criteria for lung-cancer screening. *N Engl J Med*. 2013;368(8):728–736.
32. Meza R, Jeon J, Toumazis I, et al. Evaluation of the benefits and harms of lung cancer screening with low-dose computed tomography: modeling study for the US preventive services task force. *JAMA*. 2021;325(10):988–997.
33. Ten Haaf K, Bastani M, Cao P, et al. A comparative modeling analysis of risk-based lung cancer screening strategies. *J Natl Cancer Inst*. 2020;112(5):466–479.
34. Parsons A, Daley A, Begh R, Aveyard P. Influence of smoking cessation after diagnosis of early stage lung cancer on prognosis: systematic review of observational studies with meta-analysis. *BMJ*. 2010;340:b5569.
35. Aredo JV, Luo SJ, Gardner RM, et al. Tobacco smoking and risk of second primary lung cancer. *J Thorac Oncol*. 2021;16(6):968–979.