Supplementary Materials

CONTENTS Page

Supplementary Methods

1. Model Development and Evaluation Procedure.

From the extensive epidemiological risk factor information in MEC, we considered model predictors for entry to the proposed SPLC model based on the following steps. First, we defined variables with liberal univariable P-values <0.1. Second, we included all the defined variables in a model and evaluated the model performance using three predictive performance metrics including discrimination (area under the curve; AUC), calibration, and Brier score (1). The AUC refers the discriminative ability of a model to distinguish individuals with SPLC from non-SPLC cases. Model calibration, a measure of the overall agreement between the observed outcomes and the predicted probability, was assessed by plotting a calibration curve between the observed probabilities and the expected event status obtained in groups by deciles of the predicted event probabilities. Overall prediction accuracy of the model was summarized in the Brier score that assessed the overall deviation of risk predictions estimated by models from the observed rates, of which a smaller value indicates a better model (2-4).

As parsimonious models enhance the accuracy of a model and avoid overfitting, we dropped one variable at each time using stepwise elimination methods and checked if the predictive performance improved using the three metrics (5). At each removal, the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) were calculated to score each model based on its log-likelihood and complexity (6,7). In addition, the effects of the continuous variables in MEC were evaluated in different forms (categorical vs. continuous, or linear vs. non-linear), and they were incorporated into the models using the best forms in terms of predictive performance. In particular, we examined different forms of smoking variables, which included pack-years, intensity, quit-years, and duration as well as smoking status. Then, we checked all possible pairs of interactions between the selected predictors in the model. We compared the performance of the final model with those selected by automated stepwise selection methods under backward or forward selections using BIC or AIC, which did not take into account calibration, discrimination, or predictive accuracy in model development. The final model is presented in Table 2.

After we finalized the proposed SPLC model, we used a bootstrap method (5) to obtain unbiased internal assessments of the predictive accuracy metrics by optimism correction. This method corrects for potential model overfitting by estimating the degree to which the predictive ability of a model can be expected to degrade when applied to an independent dataset.

The model was trained on 200 bootstrap samples drawn with replacement of the same size as the original data (N=5,354). Then, the bootstrapping quantifies the optimism of a prediction model between the original dataset and bootstrapped resamples to be subtracted from the original estimates for the performance metrics.

2. Decision Curve Analysis.

Decision curve analysis has been proposed to translate the effects of diagnostic and prognostic models into the clinical outcome. While the area under the curve (AUC) metric and calibration are utilized to evaluate the predictive accuracy of a model, they do not have a direct clinical interpretation and cannot tell clinical usefulness of a model among individuals varying in risk levels. Decision-analytic methods incorporate clinical consequences and evaluate whether using a risk prediction model provides larger benefits than applying other alternative strategies, such as screening everyone or screening no one.

To ascertain the value of a risk prediction model, decision curve analysis calculates the net benefit based on the concept of true positive and false-positive count among individuals underwent screening by weighting the relative harm of a false-positive and a false-negative result. A risk prediction model is considered as clinically useful when net benefit is positive, and a good model has a high net benefit.

In a standard risk model with a binary outcome, net benefit is calculated as follows (8):

$$\textbf{Net benefit} = \frac{\textbf{True positive count} - (\textbf{false positive count X Weighting factor})}{\textbf{Total number of individuals assessed for screening eligibility}}$$

Where, weighting factor, a function of threshold probability, is defined as:

$$\textbf{Weighting factor} = \frac{\textbf{Risk threshold}}{\textbf{1} - \textbf{Risk threshold}}$$

The weighting factor represents how the relative harms of false-positive results and false-negative results are valued at a given risk threshold, and thus provides the range of the threshold probability at which an individual is eligible for screening. Extensions to decision curve analysis are applicable to survival time data including competing risk (9). Decision curve also visualizes the net benefit over a range of risk thresholds, allowing one to discern whether and at which risk thresholds applying the risk model can be clinically useful.

3. Handling missing data and Multiple imputation.

Overall, the missing rate of the variables included in our data from the MEC was relatively low, mostly between 0-3.2%, except IPLC stage (6.1%) and surgery of IPLC (8.8%) (see eTable 2). Therefore, as primary analysis, we used a complete case analysis that used the subjects with complete data for the variables used for each model building process. The final prediction model with six variables (IPLC stage, IPLC histology, surgery of IPLC, prior history of cancer, the 2013 USPSTF criteria eligibility, and smoking intensity measured by cigarettes per day) included 5,354 subjects who have complete data for these six variables (see eTable 3).

For sensitivity analysis, we performed multiple imputation by chained equations to impute missing data 10 times using all of the variables included in Table 1 (10,11). For each imputed dataset, we fitted a model with the same variables included in the final prediction model and obtained a set of regression coefficients. We used Rubin's rules (12) to obtain the pooled estimates for the regression coefficients of the models fitted across the 10 imputed datasets and calculated a single predicted risk score for each study participant using the pooled estimates (eTable 5). We evaluated the model performance (discrimination, calibration, and predictive accuracy) based on the predicted risk scores calculated using the pooled regression coefficient estimates (eFigure 5).

4. Smoking data in the MEC.

10-year follow-up data for smoking information

In the Multiethnic Cohort (MEC), participants' demographic, environmental and clinical risk factors including smoking-related variables were collected from a baseline survey at enrollment between 1993 and 1996. During 2003 and 2008, a 10-year follow-up survey was administered, which captured a change in smoking status and smoking-related information. Out of 6,325 subjects included in our primary analyses, 2,044 participants submitted reassessed smoking information at 10-year follow-up. Given the goal of our study is to evaluate the effects of potential risk factors (including smoking) on the risk of SPLC using the patient data collected at the time of IPLC diagnosis, we updated smoking information for those whose follow-up data was available prior to the time of IPLC diagnosis (N=1,693, 26.8%, see eTable 1).

Projection of smoking variables to the time of IPLC diagnosis

There is a temporal gap between IPLC diagnosis and the smoking assessment (either at baseline or 10-year follow-up) in the MEC, and therefore, we conducted sensitivity analyses by using the projected smoking variables (pack-years, quit-years, duration) for evaluating the proposed prediction model. That is, among former smokers, the quit-years were projected among 3,304 subjects by adding the duration from the smoking assessment (mean of quit-year: 10.8) to diagnosis of IPLC (mean of quit-year after projection: 20.0). For current smokers, smoking duration was projected for a period from smoking assessment (mean of pack-year: 32.8) to diagnosis of IPLC among 2,914 subjects (mean of pack-year after projection: 40.3). Smoking intensity, measured by cigarettes per day (CPD) at baseline, was assumed to be constant during follow-up. Pack-years were projected with projected duration and constant smoking intensity (CPD). The eligibility to the USPSTF criteria for lung cancer screening was updated by using projected smoking information accordingly among 413 subjects.

5. Overview of the external study cohort from the PLCO.

The PLCO recruited participants aged 55 to 74 years from 1993 to 2001 to evaluate screening effectiveness in mortality reduction from cancer. Cancer and deaths were ascertained through annual follow-up of self-reports, family/relative reports and death certificates from the National Death Index. Although PLCO did not collect information on second primary malignancies following initial diagnosis, participants who were alive at trial close-out were informed of continued follow-up via state tumor registry in 2011 and could request or refuse continued

followed-up. Using state registry data from 1993 through 2014, second primary lung cancer (SPLC) status could be determined.

Therefore, our study included 2,963 ever-smoking participants in PLCO who were diagnosed with initial primary lung cancer (IPLC) between 1993 and 2014 (See Diagram 1); each IPLC case was followed up for SPLC via linkage to state registry data from 1993 to 2014. Initial tumor characteristics including stage and histology and treatment information for IPLC was collected at the time of IPLC diagnosis. In the PLCO, smoking data were collected at baseline questionnaire and updated with 10-year supplemental questionnaire (SQX) administered in 2006-2007 (N=834, 28.1%).

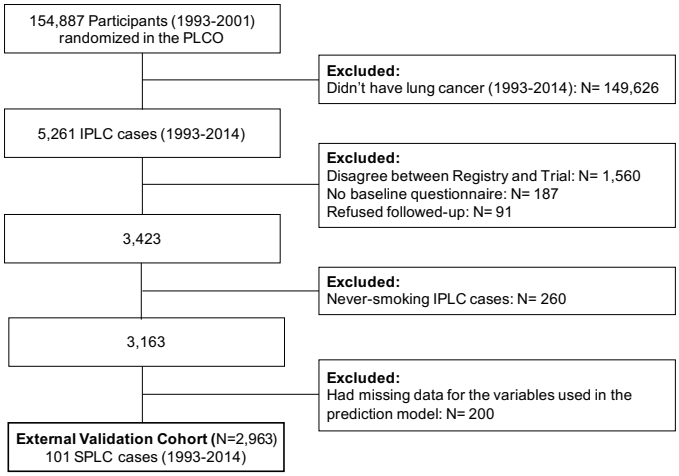


Diagram 1. External study population in the PLCO

6. Overview of the external study cohort from the NLST.

The NLST enrolled 53,454 participants aged 55 to 74 years who had a history of smoking of at least 30 pack-years, and, if former smokers, had quit within the previous 15 years during 2002 and 2004, with the purpose of determining the effect of low-dose computed tomography in lung cancer mortality reduction. NLST data included information on initial and second primary malignancies, through 2014. Our study extended follow-up for SPLC via state tumor registries, through 2014 (See Diagram 2). Data on risk factor variables including demographic (age, sex, race and education) were collected with the use of epidemiologic questionnaires administered at study entry.

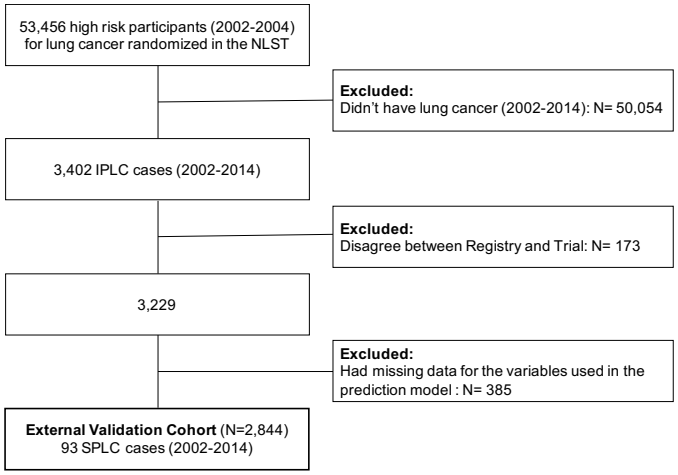


Diagram 2. External study population in the NLST

7. Predictor variables not selected in the final model.

Age at IPLC diagnosis was one of the significant predictors based on literatures, showing lower incidence of SPLC among LC survivors aged <45 or >75 years (13). In the MEC, however, age less than 45 years were not recruited; inclusion of age at IPLC diagnosis in our model did not improve predictive performance. Thus, we did not include it in the final model, but rather conducted subgroup analysis by age groups.

Smoking-related factors, such as pack-years, duration, and quit-years were not directly included in the prediction model but captured by USPSTF eligibility. Interestingly, the 2013 USPSTF variable, this summary measure of smoking duration and intensity with age eligibility (55-80 years), was a stronger indicator of a two-fold risk of SPLC among IPLC cases, than other individual smoking factors. Although smoking is found to play a significant role in SPLC development in our ongoing projects as well as current prediction study, the study subjects in SPLC analyses consist of lung cancer cases who are more likely to have a positive smoking history compared to studies study of lung cancer cases versus non-cases; moreover, over the trajectory of SPLC development from the IPLC, the carcinogenic effects of smoking on second primaries seem to be modulated (14).

The MEC is also diverse with race/ethnicity, having 20% of Asian population. However, current study did not find significant racial effect. Moreover, our proposed SPLC model is built based on ever-smokers to extend the current LC guideline and to take into account the potential etiologic differences between smoking and non-smoking LC cases, but future studies targeting SPLC among never-smokers could reveal racial/ethnic effect on SPLC risk.

Supplementary References

1. 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.
2. Zhang Z, Cortese G, Combescure C et al. Overview of model validation for survival regression model with competing risks using melanoma study data. Ann Transl Med 2018;6(16):325.
3. Gerds TA, Schumacher M. Consistent estimation of the expected Brier score in general survival models with right-censored event times. Biom J 2006;48(6):1029-1040.
4. Binder H, Allignol A, Schumacher M, Beyersmann J. Boosting for high-dimensional time-to-event data with competing risks. Bioinformatics 2009;25(7):890-896.
5. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15(4):361-387.
6. Akaike H. A new look at the statistical model identification. IEEE Transactions on Automatic Control 1974;19(6):716-723.
7. Schwarz G. Estimating the Dimension of a Model. The Annals of Statistics 1978;6(2):461-464.
8. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 2006;26(6):565-574.
9. 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.
10. Zhang Z. Multiple imputation with multivariate imputation by chained equation (MICE) package. Ann Transl Med 2016;4(2):30.
11. 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.
12. Schafer JL. Multiple imputation: a primer. Stat Methods Med Res 1999;8(1):3-15.
13. Han SS, Rivera GA, Tammemagi MC et al. Risk Stratification for Second Primary Lung Cancer. J Clin Oncol 2017;35(25):2893-2899.
14. Centers for Disease C, Prevention, National Center for Chronic Disease P et al. Publications and Reports of the Surgeon General. In How Tobacco Smoke Causes Disease: The Biology and Behavioral Basis for Smoking-Attributable Disease: A Report of the Surgeon General. Atlanta (GA): Centers for Disease Control and Prevention (US) 2010.

Supplementary Tables

Supplementary Table 1. Baseline characteristics of participants with or without 10-year follow-up smoking data in the MEC.

Variables	Updated with 10-year follow-up N=1,693 (26.8%)		Baseline data used N=4,632 (73.2%)	
Demographic Information				
Age at IPLC diagnosis				
Mean (SD)	77.8	(7.4)	72.8	(8.1)
Sex				
Female	682	(40.3)	2,529	(54.6)
Male	1,011	(59.7)	3,796	(82.0)
Race				
White	449	(26.5)	1,142	(24.7)
Japanese American	437	(25.8)	920	(19.9)
African American	338	(20.0)	1,398	(30.2)
Latino	264	(15.6)	560	(12.1)
Native Hawaiian	141	(8.3)	392	(8.5)
Others	64	(3.8)	220	(4.7)
Education				
High school	754	(44.5)	2,504	(54.1)
College	757	(44.7)	1,765	(38.1)
Postgraduate	175	(10.3)	337	(7.3)
Data missing	0	(0.0)	26	(0.6)
Smoking-related factors				
Smoking status				
Former	1,219	(72.0)	2,133	(46.0)
Current	474	(28.0)	2,499	(54.0)
Smoking intensity (Cigarettes Per Day)	17.9	(9.9)	17.6	(8.3)
Mean (SD)				
Data missing	64	(3.8)	73	(1.6)
Pack-years	35.5	(20.2)	29.4	(17.5)
Mean (SD)				
Data missing	72	(4.3)	128	(2.8)
Met the 2013 USPSTF criteria				
No	1,126	(66.5)	2,951	(63.7)
Yes	543	(32.1)	1,594	(34.4)
Data missing	24	(1.4)	87	(1.9)
Quit-years[a]				
Mean (SD)	11.6	(6.0)	10.4	(7.2)
Data missing	11	(0.6)	37	(0.8)
Clinical factors				
Prior history of cancer[b]				
No	1,205	(71.2)	3,507	(75.7)
Yes	488	(28.8)	1,125	(24.3)
BMI				
Mean (SD)	26.1	(4.6)	25.9	(4.7)
Data missing	5	(0.3)	71	(1.5)

[a]Among former smokers only. SPLC = Second primary lung cancer; IPLC = Initial primary lung cancer; SD = Standard deviation; USPSTF = United States Preventive Services Task Force; BMI = Body mass index.
[b]History of cancer, other than lung, before the time of IPLC diagnosis.

Supplementary Table 2. Missing counts and rates of variables by outcome status in the MEC.[a]

All ever-smoking cases in the MEC	Total (N=6,325)		Outcome							
			SPLC (n=145)		IPLC death (n=4,093)		Other death (n=1,049)		Censored (n=1,038)	
Age at IPLC diagnosis	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)
Sex	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)
Race	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)
Education	33	(0.5)	0	(0.0)	21	(0.5)	7	(0.7)	5	(0.5)
IPLC Stage	386	(6.1)	1	(0.7)	280	(6.8)	60	(5.7)	45	(4.3)
IPLC Histology	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)
Prior history of cancer	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)
BMI	76	(1.2)	0	(0.0)	50	(1.2)	16	(1.5)	10	(1.0)
IPLC Surgery	554	(8.8)	8	(5.5)	402	(9.8)	110	(10.5)	34	(3.3)
IPLC Radiation	169	(2.7)	0	(0.0)	121	(3.0)	20	(1.9)	28	(2.7)
IPLC Chemotherapy	302	(4.8)	1	(0.7)	207	(5.1)	57	(5.4)	37	(3.6)
Smoking status	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)
Met the 2013 USPSTF criteria	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)
Smoking intensity (Cigs/day)	137	(2.2)	1	(0.7)	86	(2.1)	27	(2.6)	23	(2.2)
Smoking pack-years	200	(3.2)	2	(1.4)	125	(3.1)	37	(3.5)	36	(3.5)
Smoking quit years	48	(0.8)	0	(0.0)	26	(0.6)	11	(1.0)	11	(1.1)

[a]SPLC = Second primary lung cancer; IPLC = Initial primary lung cancer; BMI = Body mass index; USPSTF = United States Preventive Services Task Force

Supplementary Table 3. Characteristic of IPLC patients included in the proposed SPLC model in the MEC.

Variables	Total		Outcome							
			SPLC		IPLC Death		Other Death		Censored	
	No.	(%)	No.	(%)	No.	(%)	No.	(%)	No.	(%)
Total No. of events	5,354	(100.0)	135	(100.0)	3,412	(100.0)	860	(100.0)	947	(100.0)
Follow-up time (Year)										
Mean (IQR)	2.3	(0.3-2.7)	4.3	(1.0-6.3)	1.0	(0.2-1.3)	2.2	(0.2-2.9)	6.6	(3.0-9.0)
Demographic information										
Age at IPLC diagnosis										
Mean (SD)	74.4	(8.1)	72.5	(8.2)	73.9	(8.0)	74.7	(9.7)	76.1	(8.4)
Age groups										
<55	80	(1.5)	3	(2.2)	58	(1.7)	9	(1.0)	10	(1.1)
56-60	182	(3.4)	7	(5.2)	136	(4.0)	20	(2.3)	19	(2.0)
61-70	1,239	(23.1)	39	(28.9)	803	(23.5)	201	(23.5)	196	(20.7)
71-80	2,503	(46.8)	65	(48.1)	1635	(47.9)	408	(47.4)	395	(41.7)
80+	1,350	(25.2)	21	(15.6)	780	(22.9)	222	(25.8)	327	(34.5)
Sex										
Female	2,175	(40.6)	61	(45.2)	1,326	(38.9)	331	(38.5)	457	(48.3)
Male	3,179	(59.4)	74	(54.8)	2,086	(61.1)	529	(61.5)	490	(51.7)
Race										
White	1,316	(24.6)	37	(27.4)	851	(24.9)	202	(23.5)	226	(23.9)
Japanese American	1,106	(20.7)	31	(23.0)	700	(20.5)	174	(20.2)	201	(21.2)
African American	1,558	(29.1)	37	(27.4)	1,001	(29.3)	266	(30.9)	254	(26.8)
Latino	718	(13.4)	15	(11.1)	434	(12.7)	103	(12.0)	166	(17.5)
Native Hawaiian	418	(7.8)	11	(8.1)	274	(8.0)	71	(8.3)	62	(6.5)
Others	238	(4.4)	4	(3.0)	152	(4.5)	44	(5.1)	38	(4.0)
Education										
High school	2,704	(50.5)	58	(43.0)	1,769	(51.8)	457	(53.1)	420	(44.4)
College	2,191	(40.9)	63	(46.7)	1,362	(39.9)	334	(38.8)	432	(45.6)
Postgraduate	439	(8.2)	14	(10.4)	269	(7.9)	66	(7.7)	90	(9.5)
Data missing	20	(0.4)	0	(0.0)	12	(0.4)	3	(0.3)	5	(0.5)
Tumor characteristics										
Stage of IPLC[a]										
Early stage	2,254	(42.1)	125	(92.6)	1,077	(31.6)	441	(51.3)	611	(64.5)
Advanced stage	3,100	(57.9)	10	(7.4)	2,335	(68.4)	419	(48.7)	336	(35.5)
Histology of IPLC										
Squamous cell	1,185	(22.1)	33	(24.4)	713	(20.9)	214	(24.9)	225	(23.8)
Adenocarcinoma	2,053	(38.3)	75	(55.6)	1,183	(34.7)	313	(36.4)	482	(50.9)
Large cell	163	(3.0)	7	(5.2)	105	(3.1)	29	(3.4)	21	(2.2)
Small cell	624	(11.7)	3	(2.2)	473	(13.9)	76	(8.8)	72	(7.6)
Non-small cell carcinoma, NOS	473	(8.8)	4	(3.0)	364	(10.7)	69	(8.0)	36	(3.8)
Others[b]	856	(16.0)	13	(9.6)	573	(16.8)	159	(18.5)	111	(11.7)
Smoking-related factors[c]										
Smoking status										
Former	2,801	(52.3)	69	(51.1)	1,750	(51.3)	457	(53.1)	525	(55.4)
Current	2,553	(47.7)	66	(48.9)	1,662	(48.7)	403	(46.9)	422	(44.6)
Smoking intensity (Cigarettes Per Day)										
Mean (SD)	17.6	(8.7)	19.6	(9.4)	17.8	(8.7)	17.5	(8.6)	16.6	(8.8)
Pack-years										
Mean (SD)	30.7	(18.4)	34.6	(19.3)	31.6	(18.5)	29.8	(17.8)	27.8	(18.0)
Data missing	22	(0.4)	0	(0.0)	14	(0.4)	1	(0.1)	7	(0.7)
Met the 2013 USPSTF criteria[d]										
No	3,539	(66.1)	71	(52.6)	2,172	(63.7)	587	(68.3)	709	(74.9)
Yes	1,815	(33.9)	64	(47.4)	1,240	(36.3)	273	(31.7)	238	(25.1)
Quit-years[e]										
Mean (SD)	10.9	(6.8)	9.6	(7.0)	10.5	(6.8)	11.0	(6.8)	11.9	(6.5)
Data missing	22	(0.4)	0	(0.0)	11	(0.3)	4	(0.5)	7	(0.7)
Clinical factors										
Prior history of cancer[f]										
No	3,949	(73.8)	88	(65.2)	2,597	(76.1)	577	(67.1)	687	(72.5)
Yes	1,405	(26.2)	47	(34.8)	815	(23.9)	283	(32.9)	260	(27.5)
BMI										
Mean (SD)	26.0	(4.6)	26.5	(4.7)	25.8	(4.5)	26.2	(5.2)	26.3	(4.6)
Data missing	59	(1.1)	0	(0.0)	37	(1.1)	14	(1.6)	8	(0.8)
Treatments for IPLC										
Radiotherapy										
No	3,427	(64.0)	113	(83.7)	1,994	(58.4)	631	(73.4)	689	(72.8)
Yes	1,893	(35.4)	22	(16.3)	1,397	(40.9)	226	(26.3)	248	(26.2)

Data missing	34	(0.6)	0	(0.0)	21	(0.6)	3	(0.3)	10	(1.1)
Chemotherapy										
No	3,357	(62.7)	119	(88.1)	2,007	(58.8)	606	(70.5)	625	(66.0)
Yes	1,836	(34.3)	15	(11.1)	1,300	(38.1)	218	(25.3)	303	(32.0)
Data missing	161	(3.0)	1	(0.7)	105	(3.1)	36	(4.2)	19	(2.0)
Surgery										
No	4,088	(76.4)	25	(18.5)	2,959	(86.7)	594	(69.1)	510	(53.9)
Yes	1,266	(23.6)	110	(81.5)	453	(13.3)	266	(30.9)	437	(46.1)

[a]Disease extent was defined using SEER Extent of Disease as local and regional for early stage and distant for advanced stage. SPLC = Second primary lung cancer; IPLC = Initial primary lung cancer; IQR = interquartile range; SD = Standard deviation; NOS = Not otherwise specified; USPSTF = United States Preventive Services Task Force; BMI = Body mass index.

[b]Classification of 'other' histology based on ICD-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.

[c]Smoking data were updated with available 10-year follow-up information close/prior to IPLC diagnosis for 26.8%.

[d]Aged 55-80 years, smoked ≧30 pack-years of smoking and ≦15 years since cessation

[e]Among former smokers only.

[f]History of cancer, other than lung, before the time of IPLC diagnosis.

Supplementary Table 4. Sensitivity analysis applying a Fine-Gray subdistribution hazards model to estimate the parameters of the proposed SPLC model in the MEC.[a]

Factors	No.	A subdistribution hazard model	
		HR (95% CI)	P-value
Histology of IPLC			
Squamous cell	1,185	Ref	
Large cell	2,053	1.55 (0.69-3.49)	0.290
Adenocarcinoma	163	1.28 (0.83-1.96)	0.270
Small cell	624	0.77 (0.23-2.55)	0.670
Non-small cell carcinoma, NOS	473	0.93 (0.32-2.71)	0.900
Other	856	1.08 (0.58-2.00)	0.810
Prior history of cancer[b]			
No	3,949	Ref	
Yes	1,405	1.33 (0.93-1.91)	0.120
Met the 2013 USPSTF criteria[c]			
No	3,539	Ref	
Yes	1,815	1.73 (1.15-2.61)	0.009
Smoking intensity (Cigs/day)	5,354	1.01 (0.99-1.04)	0.350
Surgery for IPLC			
No	2,525	Ref	
Yes	3,414	4.99 (2.93-8.51)	$3.4X10^{-9}$
Stage of IPLC			
Early stage	2,254	Ref	
Advanced stage	3,100	0.27 (0.12-0.60)	0.001
Stage of IPLC x Met the 2013 USPSTF criteria[c]		0.27 (0.05-1.32)	0.100

[a]Based-on N=5,354 subjects in the MEC data. The hazards of the factors included in the proposed SPLC model were re-fitted using Fine-Gray subdistribution proportional hazard regression. HR = Subdistribution hazard ratio; 95% CI = 95% Confidence interval; SPLC = second primary lung cancer; IPLC = Initial primary lung cancer; NOS = Not otherwise specified; USPSTF = United States Preventive Services Task Force.

[b]History of cancer, other than lung, before the time of IPLC diagnosis.

[c]Aged 55-80 years, smoked ≧30 pack-years of smoking and ≦15 years since cessation.

Supplementary Table 5. Sensitivity Analysis using multiple imputation to estimate the parameters of the proposed SPLC model in the MEC.[a]

Fators.	No.	A pooled cause-specific hazard model	
		HR (95% CI)	P-value
Histology of IPLC			
Squamous cell	1,369	Ref	
Large cell	194	2.09 (1.00 to 4.37)	0.050
Adenocarcinoma	2,348	1.17 (0.78 to 1.75)	0.441
Small cell	729	0.68 (0.20 to 2.27)	0.527
Non-small cell carcinoma, NOS	504	0.97 (0.33 to 2.82)	0.958
Other	1,190	0.87 (0.46 to 1.66)	0.673
Prior history of cancer[b]			
No	4,712	Ref	
Yes	1,613	1.36 (0.96 to 1.93)	0.088
Met the 2013 USPSTF criteria[c]			
No	4,136	Ref	
Yes	2,189	1.79 (1.20 to 2.67)	0.004
Smoking intensity (Cigs/day)	6,325	1.01 (0.99 to 1.04)	0.272
Surgery for IPLC			
No	4,877	Ref	
Yes	1,448	2.20 (1.28 to 3.76)	0.004
Stage of IPLC			
Early stage	2,647	Ref	
Advanced stage	3,678	0.47 (0.21 to 1.05)	0.067
Stage of IPLC x Met the 2013 USPSTF criteria[c]		0.28 (0.06 to 1.40)	0.122

[a]Based-on 10 completed datasets (N=6,325X10) by multiple imputation in the MEC. Hazards of the factors included in the proposed SPLC model were re-fitted in each of imputed datasets and pooled using Rubin's rules. HR = Hazard ratio; 95% CI = 95% Confidence interval; IPLC = Initial primary lung cancer; NOS = Not otherwise specified; USPSTF= United States Preventive Services Task Force.

[b]History of cancer, other than lung, before the time of IPLC diagnosis.

[c]Aged 55-80 years, smoked ≧30 pack-years of smoking and ≦15 years since cessation

Supplementary Table 6. Sensitivity Analysis using the projected smoking data to estimate the parameters of the proposed SPLC model in the MEC.[a]

Factors	No.	A cause-specific hazard model	
		HR (95% CI)	P-value
Histology of IPLC			
Squamous cell	1,185	Ref	
Large cell	2,053	2.00 (0.88 to 4.56)	0.097
Adenocarcinoma	163	1.16 (0.77 to 1.76)	0.485
Small cell	624	0.81 (0.24 to 2.75)	0.735
Non-small cell carcinoma, NOS	473	0.89 (0.31 to 2.60)	0.837
Other	856	0.95 (0.50 to 1.82)	0.885
Prior history of cancer[b]			
No	3,949	Ref	
Yes	1,405	1.42 (0.99 to 2.02)	0.057
Met the 2013 USPSTF criteria[c]			
No	3,539	Ref	
Yes	1,815	1.58 (1.07 to 2.34)	0.021
Smoking intensity (Cigs/day)	5,354	1.02 (1.00 to 1.04)	0.121
Surgery for IPLC			
No	2,525	Ref	
Yes	3,414	2.15 (1.26 to 3.69)	0.005
Stage of IPLC			
Early stage	2,254	Ref	
Advanced stage	3,100	0.49 (0.22 to 1.10)	0.085
Stage of IPLC x Met the 2013 USPSTF criteria[c]		0.26 (0.05 to 1.28)	0.098

[a]Based-on N=5,354 subject in the MEC data. Hazards of the factors included in the proposed SPLC prediction model were re-fitted using updated USPSTF criteria from projected smoking data. HR = Hazard ratio; 95% CI = 95% Confidence interval; IPLC = Initial primary lung cancer; NOS = Not otherwise specified; USPSTF = United States Preventive Services Task Force.

[b]History of cancer, other than lung, before the time of IPLC diagnosis.

[c]Aged 55-80 years, smoked ≧30 pack-years of smoking and ≦15 years since cessation

Supplementary Table 7. Characteristic of never-smoking patients of lung cancer in the MEC.

Variable	Total		Outcome							
			SPLC		IPLC Death		Other Death		Censored	
	No.	(%)	No.	(%)	No.	(%)	No.	(%)	No.	(%)
Total No. of events	740	(100.0)	17	(100.0)	389	(100.0)	121	(100.0)	213	(100.0)
Follow-up time (Year)										
Mean (IQR)	3.2	(0.4-4.3)	3.7	(1.3-5.6)	1.4	(0.3-1.8)	2.1	(0.1-3.0)	6.9	(2.8-9.3)
Demographic information										
Age at IPLC diagnosis										
Mean (SD)	75.7	(8.6)	73.4	(7.9)	75.2	(8.4)	77.3	(8.0)	76.0	(9.1)
Age groups										
<55	13	(1.8)	0	(0.0)	8	(2.1)	2	(1.7)	3	(1.4)
56-60	23	(3.1)	1	(5.9)	14	(3.6)	1	(0.8)	7	(3.3)
61-70	134	(18.1)	3	(17.6)	63	(16.2)	20	(16.5)	48	(22.5)
71-80	326	(44.1)	11	(64.7)	187	(48.1)	48	(39.7)	80	(37.6)
80+	244	(33.0)	2	(11.8)	117	(30.1)	50	(41.3)	75	(35.2)
Sex										
Female	545	(73.6)	13	(76.5)	279	(71.7)	78	(64.5)	175	(82.2)
Male	195	(26.4)	4	(23.5)	110	(28.3)	43	(35.5)	38	(17.8)
Race										
White	107	(14.5)	3	(17.6)	47	(12.1)	21	(17.4)	36	(16.9)
Japanese American	232	(31.4)	5	(29.4)	132	(33.9)	40	(33.1)	55	(25.8)
African American	113	(15.3)	0	(0.0)	67	(17.2)	22	(18.2)	24	(11.3)
Latino	174	(23.5)	5	(29.4)	81	(20.8)	25	(20.7)	63	(29.6)
Native Hawaiian	34	(4.6)	1	(5.9)	18	(4.6)	6	(5.0)	9	(4.2)
Others	80	(10.8)	3	(17.6)	44	(11.3)	7	(5.8)	26	(12.2)
Education										
High school	339	(45.8)	7	(41.2)	183	(47.0)	61	(50.4)	88	(41.3)
College	293	(39.6)	9	(52.9)	143	(36.8)	49	(40.5)	92	(43.2)
Postgraduate	100	(13.5)	1	(5.9)	60	(15.4)	10	(8.3)	29	(13.6)
Data missing	8	(1.1)	0	(0.0)	3	(0.8)	1	(0.8)	4	(1.9)
Tumor characteristics										
Stage of IPLC[a]										
Early stage	307	(41.5)	12	(70.6)	104	(26.7)	50	(41.3)	141	(66.2)
Advanced stage	433	(58.5)	5	(29.4)	285	(73.3)	71	(58.7)	72	(33.8)
Histology of IPLC										
Squamous cell	48	(6.5)	1	(5.9)	28	(7.2)	8	(6.6)	11	(5.2)
Adenocarcinoma	475	(64.2)	13	(76.5)	248	(63.8)	66	(54.5)	148	(69.5)
Large cell	11	(1.5)	0	(0.0)	7	(1.8)	1	(0.8)	3	(1.4)
Small cell	13	(1.8)	0	(0.0)	8	(2.1)	5	(4.1)	0	(0.0)
Non-small cell carcinoma, NOS	57	(7.7)	0	(0.0)	37	(9.5)	13	(10.7)	7	(3.3)
Others[b]	136	(18.4)	3	(17.6)	61	(15.7)	28	(23.1)	44	(20.7)
Clinical factors										
Prior history of cancer[c]										
No	546	(73.8)	13	(76.5)	307	(78.9)	70	(57.9)	156	(73.2)
Yes	194	(26.2)	4	(23.5)	82	(21.1)	51	(42.1)	57	(26.8)
BMI										
Mean (SD)	25.7	(4.9)	23.7	(3.7)	25.6	(4.9)	26.4	(4.9)	25.6	(4.8)
Data missing	8	(1.1)	0	(0.0)	4	(1.0)	4	(3.3)	0	(0.0)
Treatments for IPLC										
Radiotherapy										
No	538	(72.7)	16	(94.1)	251	(64.5)	94	(77.7)	177	(83.1)
Yes	199	(26.9)	1	(5.9)	137	(35.2)	26	(21.5)	35	(16.4)
Data missing	3	(0.4)	0	(0.0)	1	(0.3)	1	(0.8)	1	(0.5)
Chemotherapy										
No	477	(64.5)	12	(70.6)	219	(56.3)	91	(75.2)	155	(72.8)
Yes	249	(33.6)	5	(29.4)	164	(42.2)	26	(21.5)	54	(25.4)
Data missing	14	(1.9)	0	(0.0)	6	(1.5)	4	(3.3)	4	(1.9)
Surgery										
No	507	(68.5)	6	(35.3)	325	(83.5)	82	(67.8)	94	(44.1)
Yes	233	(31.5)	11	(64.7)	64	(16.5)	39	(32.2)	119	(55.9)

[a]Disease extent was defined using SEER Extent of Disease as local and regional for early stage and distant for advanced stage. SPLC = Second primary lung cancer; IPLC = Initial primary lung cancer; IQR = interquartile range; SD = Standard deviation; NOS = Not otherwise specified; BMI = Body mass index.

[b]Classification of 'other' histology based on ICD-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.

[c]History of cancer, other than lung, before the time of IPLC diagnosis.

Supplementary Table 8. Sensitivity Analysis using data excluding 2,098 patients who died or were lost to follow-up within 6 months after IPLC diagnosis to estimate the parameters of the SPLC prediction model in the MEC.[a]

Factors	No.	A cause-specific hazard model	
		HR (95% CI)	P-value
Histology of IPLC			
Squamous cell	767	Ref	
Large cell	87	2.63 (1.13 to 6.09)	0.023
Adenocarcinoma	1,386	0.96 (0.60 to 1.51)	0.854
Small cell	343	1.17 (0.33 to 4.08)	0.804
Non-small cell carcinoma, NOS	258	0.57 (0.13 to 2.49)	0.459
Other	414	0.97 (0.47 to 1.96)	0.932
Prior history of cancer[b]			
No	2,353	Ref	
Yes	902	1.77 (1.19 to 2.63)	0.004
Met the 2013 USPSTF criteria[c]			
No	2,173	Ref	
Yes	1,082	1.77 (1.11 to 2.81)	0.015
Smoking intensity (Cigs/day)	3,256	1.01 (0.98 to 1.03)	0.380
Surgery for IPLC			
No	2,112	Ref	
Yes	1,143	2.17 (1.12 to 4.16)	0.020
Stage of IPLC			
Early stage	1,850	Ref	
Advanced stage	1,405	0.76 (0.29 to 1.93)	0.567
Stage of IPLC x Met the 2013 USPSTF criteria[c]		0.39 (0.07 to 2.06)	0.272

[a]Based-on N=3,256 subject (107 SPLC cases) in the MEC data. Hazards of the factors in the proposed SPLC prediction model were re-fitted using data excluding 2,098 patients who died or were lost to follow-up within 6 months after IPLC diagnosis. HR = Hazard ratio; 95% CI = 95% Confidence interval; IPLC = Initial primary lung cancer; NOS = Not otherwise specified; USPSTF = United States Preventive Services Task Force.
[b]History of cancer, other than lung, before the time of IPLC diagnosis.
[c]Aged 55-80 years, smoked ≧30 pack-years of smoking and ≦15 years since cessation.

Supplementary Table 9. Characteristics of external study population from the PLCO.

Variable	Total		Outcome							
			SPLC		IPLC Death		Other Death		Censored	
	No.	(%)	No.	(%)	No.	(%)	No.	(%)	No.	(%)
Total No. of events	2,963	(100.0)	101	(100.0)	2,182	(100.0)	320	(100.0)	360	(100.0)
Follow-up time (Year)										
Mean (SD)	2.5	(3.5)	5.5	(3.8)	1.2	(1.6)	4.7	(4.0)	7.3	(4.9)
Demographic information										
Age at IPLC diagnosis										
Mean (SD)	71.5	(6.4)	69.1	(6.1)	71.6	(6.3)	72.1	(6.6)	71.3	(6.9)
Age groups										
55-59	99	(3.3)	7	(6.9)	68	(3.1)	10	(3.1)	14	(3.9)
60-69	1,036	(35.0)	49	(48.5)	754	(34.6)	97	(30.3)	136	(37.8)
70-79	1,508	(50.9)	40	(39.6)	1,139	(52.2)	168	(52.5)	161	(44.7)
80+	320	(10.8)	5	(5.0)	221	(10.1)	45	(14.1)	49	(13.6)
Sex										
Female	1,132	(38.2)	39	(38.6)	786	(36.0)	108	(33.8)	199	(55.3)
Male	1,831	(61.8)	62	(61.4)	1,396	(64.0)	212	(66.3)	161	(44.7)
Race										
White	2,620	(88.4)	99	(97.0)	1,920	(88.0)	287	(89.7)	315	(87.5)
Asian	70	(2.4)	0	(0.0)	53	(2.4)	7	(2.2)	10	(2.8)
African American	207	(7.0)	2	(2.0)	159	(7.3)	17	(5.3)	29	(8.1)
Latino	38	(1.3)	0	(0.0)	27	(1.2)	5	(1.6)	6	(1.7)
Pacific Islander	20	(0.7)	1	(1.0)	17	(0.8)	2	(0.6)	0	(0.0)
American Indian	8	(0.3)	0	(0.0)	6	(0.3)	2	(0.6)	0	(0.0)
Education										
High school	1,565	(52.8)	60	(59.4)	1,169	(53.6)	160	(50.0)	176	(48.9)
College	1,109	(37.4)	32	(31.7)	802	(36.8)	130	(40.6)	145	(40.3)
Postgraduate	285	(9.6)	9	(8.9)	208	(9.5)	29	(9.1)	39	(10.8)
Data missing	4	(0.1)	0	(0.0)	3	(0.1)	1	(0.3)	0	(0.0)
Tumor characteristics										
Stage of IPLC[a]										
Early stage	1,625	(54.8)	92	(91.1)	925	(42.4)	272	(85.0)	336	(93.3)
Advanced stage	1,338	(45.2)	9	(8.9)	1,257	(57.6)	48	(15.0)	24	(6.7)
SEER Stage at IPLC diagnosis										
Local	744	(25.1)	61	(60.4)	270	(12.4)	177	(55.3)	236	(65.6)
Regional	881	(29.7)	31	(30.7)	655	(30.0)	95	(29.7)	100	(27.8)
Distant	1,338	(45.2)	9	(8.9)	1,257	(57.6)	48	(15.0)	24	(6.7)
Histology of IPLC										
Squamous cell	668	(22.5)	28	(27.7)	437	(20.0)	95	(29.7)	108	(30.0)
Adenocarcinoma	1,256	(42.4)	53	(52.5)	862	(39.5)	133	(41.6)	208	(57.8)
Large cell	117	(3.9)	4	(4.0)	85	(3.9)	20	(6.3)	8	(2.2)
Small cell	446	(15.1)	5	(5.0)	412	(18.9)	16	(5.0)	13	(3.6)
Others[b]	476	(16.1)	11	(10.8)	386	(17.7)	56	(17.5)	23	(6.4)
Smoking-related factors[c]										
Smoking status										
Former	1,744	(58.9)	53	(52.5)	1,281	(58.7)	180	(56.3)	230	(63.9)
Current	1,219	(41.1)	48	(47.5)	901	(41.3)	140	(43.8)	130	(36.1)
Smoking intensity (Cigarettes per day)										
Mean (SD)	24.9	(12.1)	26.4	(13.7)	25.0	(12.1)	26.1	(12.5)	23.3	(11.3)
Pack-years										
Mean (SD)	57.8	(32.8)	62.8	(34.4)	58.2	(32.8)	61.5	(34.3)	50.3	(29.3)
Met USPSTF criteria[d]										
No	1,266	(42.7)	30	(29.7)	913	(41.8)	128	(40.0)	195	(54.2)
Yes	1,697	(57.3)	71	(70.3)	1,269	(58.2)	192	(60.0)	165	(45.8)
Quit-years										
Mean (SD)	11.4	(13.4)	8.8	(12.2)	11.1	(13.3)	11.2	(13.4)	13.7	(14.1)
Clinical factors										
Prior history of cancer[e]										
No	2,735	(92.3)	94	(93.1)	2,009	(92.1)	297	(92.8)	335	(93.1)
Yes	228	(7.7)	7	(6.9)	173	(7.9)	23	(7.2)	25	(6.9)
BMI										
Mean (SD)	26.6	(4.5)	26.3	(3.6)	26.6	(4.5)	27.0	(4.8)	26.3	(4.5)
Data missing	38	(1.3)	1	(1.0)	26	(1.2)	3	(0.9)	8	(2.2)
Treatments for IPLC										
Radiotherapy										
No	1,754	(59.2)	77	(76.2)	1,162	(53.3)	230	(71.9)	285	(79.2)
Yes	1,209	(40.8)	24	(23.8)	1,020	(46.7)	90	(28.1)	75	(20.8)
Chemotherapy										

No	1,444	(48.7)	73	(72.3)	895	(41.0)	232	(72.5)	244	(67.8)
Yes	1,519	(51.3)	28	(27.7)	1,287	(59.0)	88	(27.5)	116	(32.2)
Surgery										
No	2,018	(68.1)	17	(16.8)	1,788	(81.9)	127	(39.7)	86	(23.9)
Yes	945	(31.9)	84	(83.2)	394	(18.1)	193	(60.3)	274	(76.1)

[a]Disease extent was defined using SEER Extent of Disease as local and regional for early and distant for advanced stage. PLCO = Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; SPLC = Second primary lung cancer; IPLC = Initial primary lung cancer; SD = Standard deviation; USPSTF = United States Preventive Services Task Force; BMI = Body mass index.

[b]Classification of 'other' histology based on ICD-O-3 codes including 8000, 8010, 8012, 8020, 8022, 8031, 8032, 8033, 8246 and 8560 in the PLCO; all confirmed lung cancer cases.

[c]Smoking data closest to initial diagnosis were extracted (27.9%).

[d]Aged 55-80 years, smoked ≥30 pack-years of smoking and ≤15 years since cessation.

[e]History of cancer, other than lung, before the time of IPLC diagnosis.

Supplementary Table 10. Characteristics of external study population from the NLST.

Variable	Total		Outcome							
			SPLC		IPLC Death		Other Death		Censored	
	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)
Total No. of events	2,844	(100.0)	93	(100.0)	1,498	(100.0)	239	(100.0)	1,014	(100.0)
Follow-up time (Year)										
Mean (SD)	2.6	(3.0)	4.5	(2.7)	1.2	(1.5)	3.0	(2.9)	4.4	(3.6)
Demographic information										
Age at IPLC diagnosis										
Mean (SD)	68.4	(6.1)	65.5	(5.9)	68.5	(6.1)	68.5	(6.2)	68.7	(6.1)
Age groups										
55-59	202	(7.1)	17	(18.3)	103	(6.9)	17	(7.1)	65	(6.4)
60-69	1,433	(50.4)	51	(54.8)	750	(50.1)	118	(49.4)	514	(50.7)
70-79	1,102	(38.7)	24	(25.8)	591	(39.5)	97	(40.6)	390	(38.5)
80+	107	(3.8)	1	(1.1)	54	(3.6)	7	(2.9)	45	(4.4)
Sex										
Female	1,202	(42.3)	43	(46.2)	566	(37.8)	97	(40.6)	496	(48.9)
Male	1,642	(57.7)	50	(53.8)	932	(62.2)	142	(59.4)	518	(51.1)
Race										
White	2,578	(90.6)	78	(83.9)	1,353	(90.3)	222	(92.9)	925	(91.2)
Asian	45	(1.6)	4	(4.3)	23	(1.5)	0	(0.0)	18	(1.8)
African American	141	(5.0)	9	(9.7)	75	(5.0)	14	(5.9)	43	(4.2)
Latino	34	(1.2)	1	(1.1)	15	(1.0)	1	(0.4)	17	(1.7)
Pacific Islander	6	(0.2)	0	(0.0)	4	(0.3)	0	(0.0)	2	(0.2)
American Indian	9	(0.3)	0	(0.0)	8	(0.5)	0	(0.0)	1	(0.1)
Missing	31	(1.1)	1	(1.1)	20	(1.3)	2	(0.8)	8	(0.8)
Education										
High school	1,510	(53.1)	43	(46.2)	823	(54.9)	120	(50.2)	524	(51.7)
College	1,007	(35.4)	37	(39.8)	518	(34.6)	92	(38.5)	360	(35.5)
Postgraduate	276	(9.7)	11	(11.8)	133	(8.9)	22	(9.2)	110	(10.8)
Data missing	51	(1.8)	2	(2.2)	24	(1.6)	5	(2.1)	20	(2.0)
Tumor characteristics										
Stage of IPLC[a]										
Early stage	1,761	(61.6)	88	(94.6)	610	(40.7)	180	(75.3)	873	(86.1)
Advanced stage	1,093	(38.4)	5	(5.4)	888	(59.3)	59	(24.7)	141	(13.9)
SEER Stage at IPLC diagnosis										
Local	942	(33.1)	59	(63.4)	159	(10.6)	114	(47.7)	610	(60.2)
Regional	809	(28.4)	29	(31.2)	451	(30.1)	66	(27.6)	263	(25.9)
Distant	1,093	(38.4)	5	(5.4)	888	(59.3)	59	(24.7)	141	(13.9)
Histology of IPLC										
Squamous cell	650	(22.9)	25	(26.9)	300	(20.0)	66	(27.6)	259	(25.5)
Adenocarcinoma	1,173	(41.2)	44	(47.3)	519	(34.6)	92	(38.5)	518	(51.1)
Large cell	95	(3.3)	3	(3.2)	60	(4.0)	10	(4.2)	22	(2.2)
Small cell	455	(16.0)	6	(6.5)	345	(23.0)	19	(7.9)	85	(8.4)
Non-small cell carcinoma, NOS	289	(10.2)	10	(10.8)	175	(11.7)	29	(12.1)	75	(7.4)
Others[b]	182	(6.4)	5	(5.4)	99	(6.6)	23	(9.6)	55	(5.4)
Smoking-related factors										
Smoking status										
Former	1,098	(38.6)	35	(37.6)	532	(35.5)	92	(38.5)	439	(43.3)
Current	1,746	(61.4)	58	(62.4)	966	(64.5)	147	(61.5)	575	(56.7)
Smoking intensity (Cigarettes per day)										
Mean (SD)	29.2	(11.5)	30.4	(10.5)	29.2	(11.5)	29.8	(12.0)	28.8	(11.3)
Pack-years										
Mean (SD)	63.5	(26.2)	64.7	(22.3)	64.3	(26.5)	65.7	(26.6)	61.7	(25.9)
Met USPSTF criteria[c]										
No	332	(11.7)	3	(3.2)	165	(11.0)	19	(7.9)	145	(14.3)
Yes	2,512	(88.3)	90	(96.8)	1,333	(89.0)	220	(92.1)	869	(85.7)
Quit-years										
Mean (SD)	4.4	(6.7)	3.5	(5.8)	4.0	(6.4)	4.2	(6.4)	5.1	(7.1)
Clinical factors										
Prior history of cancer[d]										
No	2,671	(93.9)	87	(93.5)	1,414	(94.4)	220	(92.1)	950	(93.7)
Yes	173	(6.1)	6	(6.5)	84	(5.6)	19	(7.9)	64	(6.3)
BMI										
Mean (SD)	26.9	(4.8)	27.8	(6.1)	27.0	(4.8)	26.5	(4.5)	26.8	(4.6)
Data missing	14	(0.5)	1	(1.1)	8	(0.5)	0	(0.0)	5	(0.5)
Treatments for IPLC										
Radiotherapy										

No	1,792	(63.0)	72	(77.4)	816	(54.5)	173	(72.4)	731	(72.1)
Yes	1,037	(36.5)	21	(22.6)	677	(45.2)	64	(26.8)	275	(27.1)
Data missing	15	(0.5)	0	(0.0)	5	(0.3)	2	(0.8)	8	(0.8)
Chemotherapy										
No	1,414	(49.7)	60	(64.5)	541	(36.1)	159	(66.5)	654	(64.5)
Yes	1,406	(49.4)	33	(35.5)	943	(63.0)	78	(32.6)	352	(34.7)
Data missing	24	(0.8)	0	(0.0)	14	(0.9)	2	(0.8)	8	(0.8)
Surgery										
No	1,632	(57.4)	15	(16.1)	1,172	(78.2)	108	(45.2)	337	(33.2)
Yes	1,212	(42.6)	78	(83.9)	326	(21.8)	131	(54.8)	677	(66.8)

[a]Disease extent was defined using SEER Extent of Disease as local and regional for early and distant for advanced stage. NLST = National Lung Screening Trial; SPLC = Second primary lung cancer; IPLC = Initial primary lung cancer; SD = Standard deviation; USPSTF = United States Preventive Services Task Force; BMI = Body mass index.

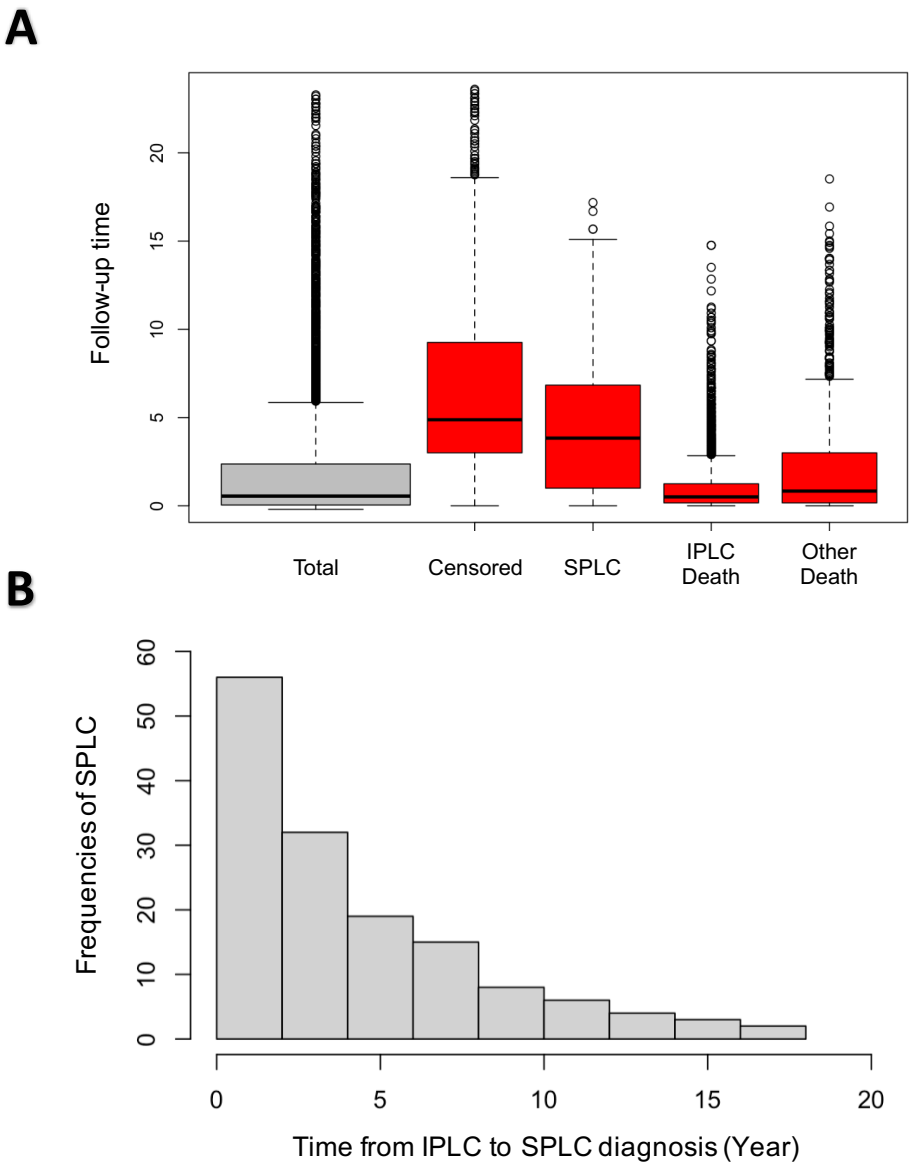
[b]Classification of 'other' histology based on ICD-O-3 codes including 8000, 8001, 8010, 8022, 8032, 8033, 8230, 8240, 8249, 8560 and 8980 in the NLST; all confirmed lung cancer cases.

[c]Aged 55-80 years, smoked ≥30 pack-years of smoking and ≤15 years since cessation.

[d]History of cancer, other than lung, before the time of IPLC diagnosis.

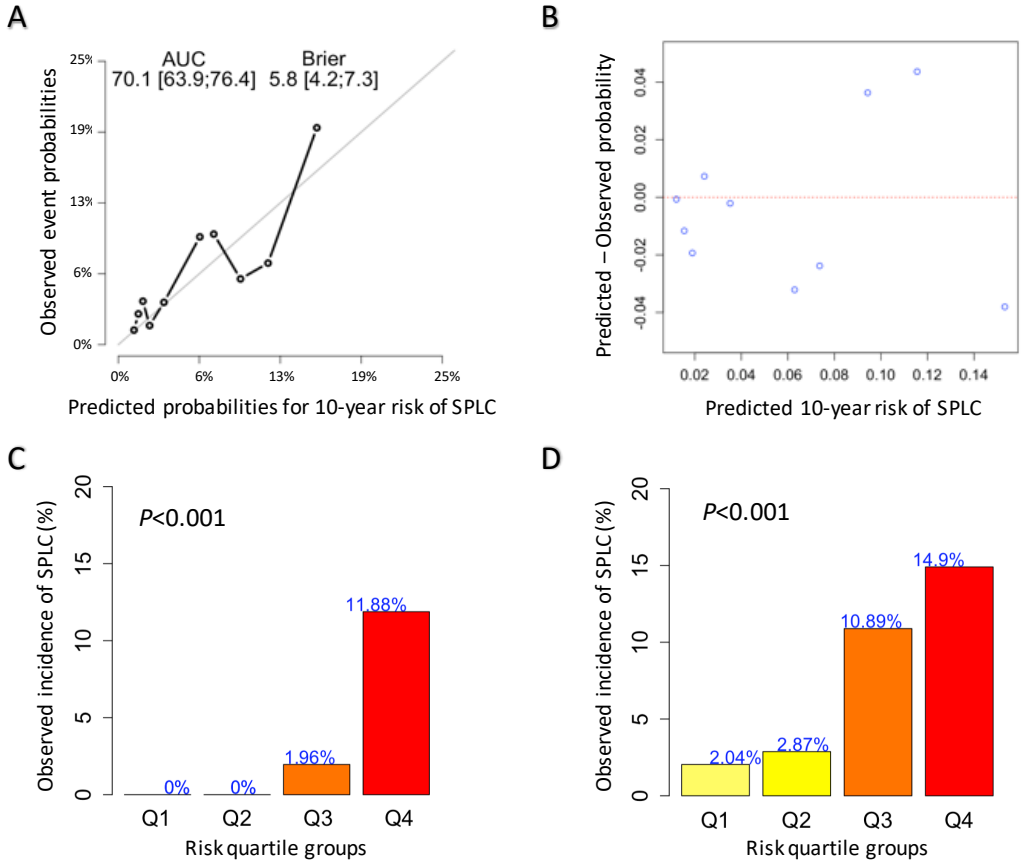
Supplementary Figures

Supplementary Figure 1. Box plot for follow-up time stratified by outcome status and the distribution of time from IPLC to SPLC diagnosis for SPLC cases (N=145) in the MEC.
(A) The y-axis shows the quartiles of follow-up time in years. The mean follow-up time for total study population was 2.2 years, and by outcomes was 4.6 years for SPLC, 1.0 year for IPLC death, 2.2 year for other death and 6.7 years for censored cases. (B) The y-axis shows the frequencies of SPLC, and x-axis is the time from IPLC to SPLC diagnosis in years, and the x-axis shows the time from IPLC to SPLC diagnosis. Abbreviations. IPLC, initial primary of lung cancer; SPLC, second primary lung cancer; MEC, the Multiethnic Cohort Study.

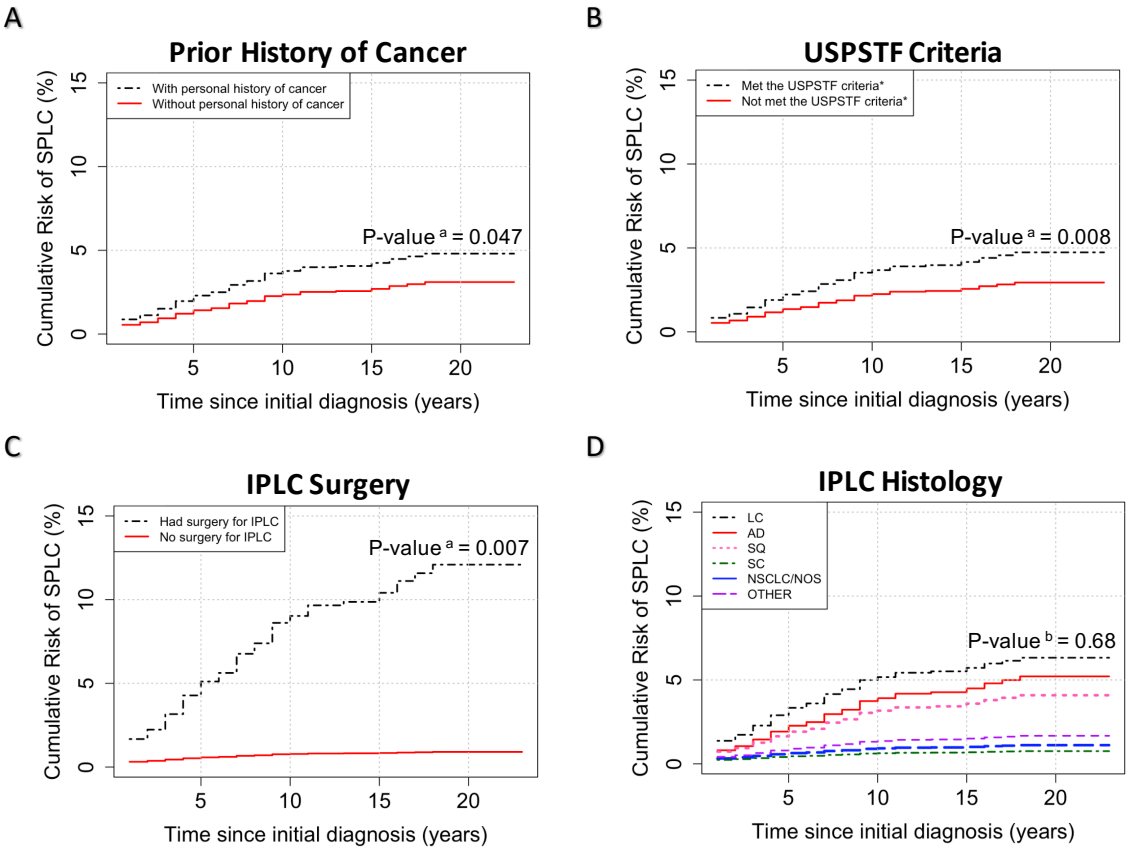


Supplementary Figure 2. Sensitivity analysis for evaluating the performance of the proposed SPLC model among early-stage IPLC patients in the MEC.

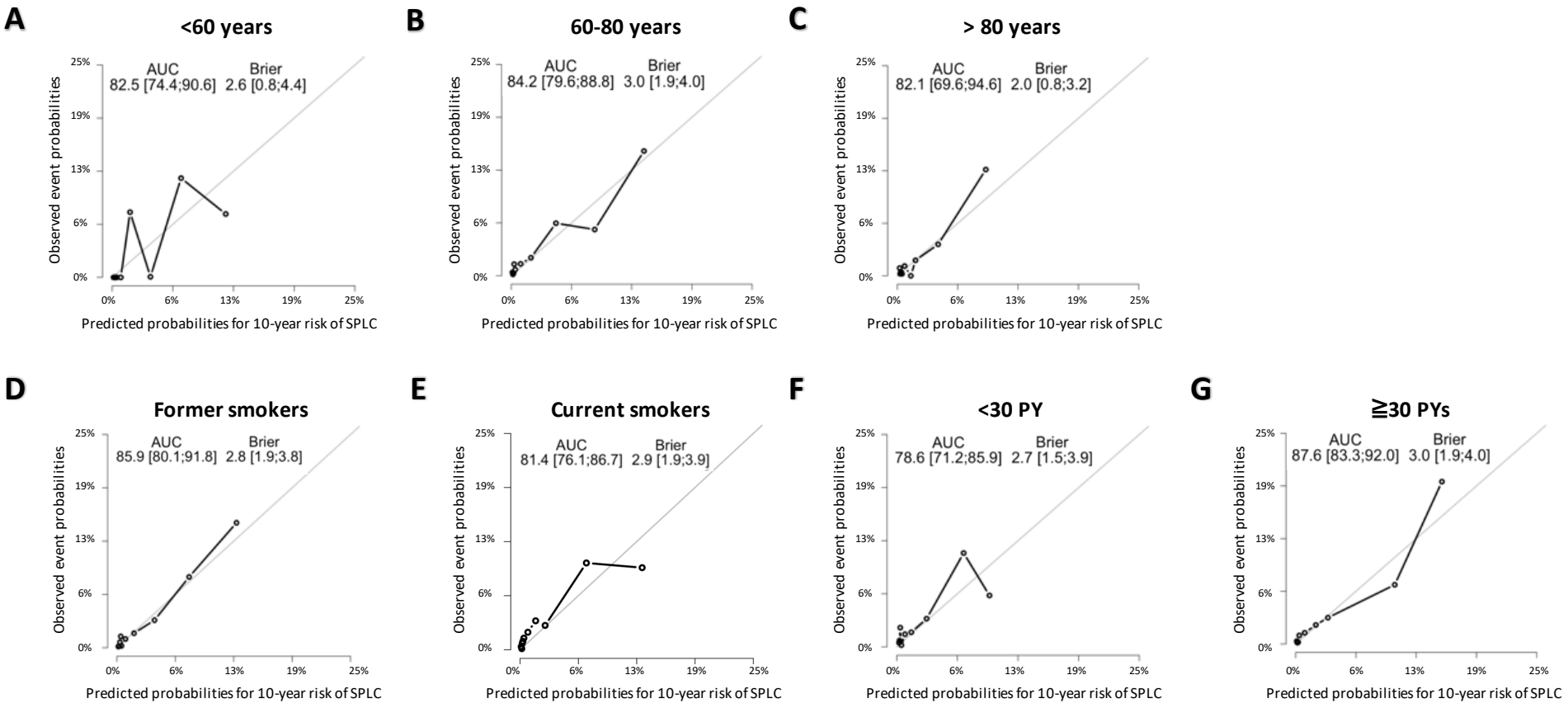
(A) Calibration plot with discriminative performance (area under the curve; AUC) and prediction accuracy (Brier score) of the proposed SPLC prediction model (see **Table 2**) among early-stage IPLC patients. (B) Calibration error of mean difference between observed and predicted 10-year SPLC probabilities by risk deciles. (C and D) Cumulative 10-year incidence of SPLC by risk-quartile groups of the estimated risk using the proposed risk prediction model. The four risk-quartile groups were divided based on the following 10-year SPLC risk thresholds obtained from the development cohort (N=5,354) in (C): Q1: r ≦0.3%, Q2: 0.3%< r ≦0.5%, Q3: 0.5%< r ≦2.8% and Q4: r >2.8%; and, obtained from early-stage IPLC patients (N=2,254) in (D): Q1: r ≦1.9%, Q2: 1.9%< r ≦5.0%, Q3: 5.0%< r ≦9.3% and Q4: r >9.3%. The risks across different groups were compared and tested using the method by Gray test. Abbreviations. IPLC, Initial primary lung cancer; SPLC, second primary lung cancer; MEC, the Multiethnic Cohort Study; AUC, Area Under the Curve.



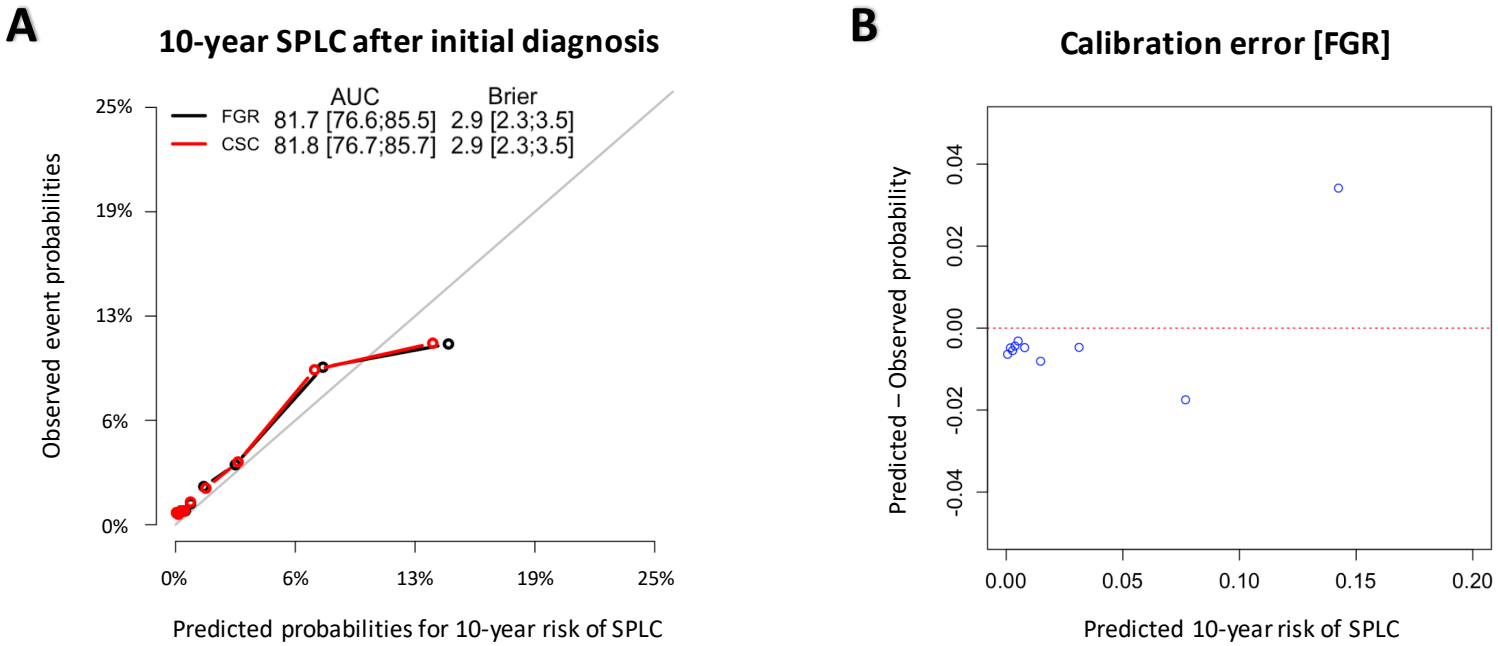
Supplementary Figure 3. Cumulative risk of SPLC among IPLC patients by key predictors included in the proposed SPLC model in MEC. (A) By prior history of cancer, (B) by eligibility to the USPSTF criteria*, (C) by status of surgery for IPLC and (D) by IPLC histologic subtype. [a] P-value obtained from the cause-specific hazard model (**Table 2**). [b] Global p-value using a likelihood-ratio test for jointly testing the effects of the multiple categories of the histologic subtypes. * Aged 55-80 years, smoked ≥30 PYs of smoking and ≤15 years since cessation. Abbreviations. IPLC, Initial primary lung cancer; SPLC, second primary lung cancer; LC, large cell; AD, adenocarcinoma; SQ, squamous cell; SC, small cell; NSCLC/NOS, non-small-cell lung cancer/not otherwise specified; USPSTF, United States Preventive Services Task Force; MEC, the Multiethnic Cohort Study.



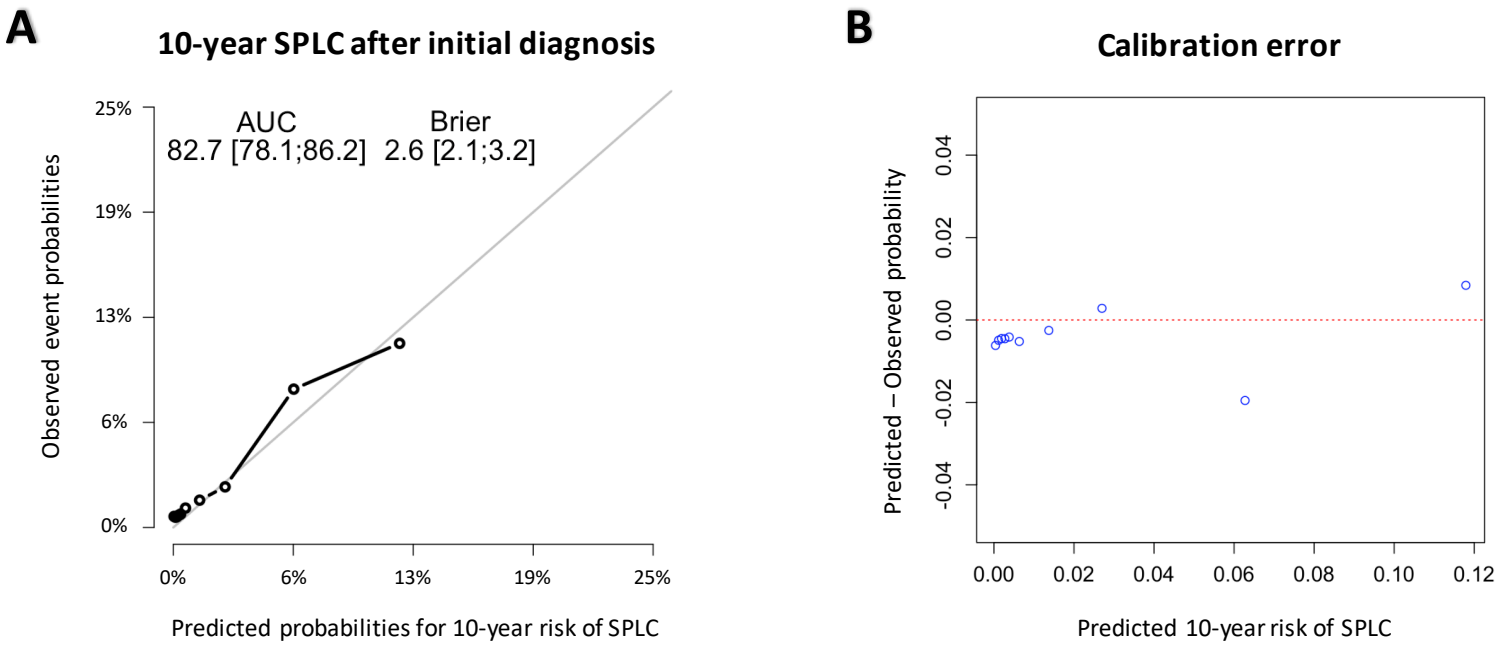
Supplementary Figure 4. Sensitivity analysis for evaluating the performance of the proposed SPLC model by subgroups in the MEC. (A to C) Calibration plots with discriminative performance (area under the curve; AUC) and prediction accuracy (Brier score) of the proposed SPLC prediction model (see **Table 2**) in subgroups of age at IPLC diagnosis (<60, 60-80, >80 years), (D,E) by smoking status (former, current), and (F, G) by pack-year threshold (<30, ≧30 Pack-years), respectively. Abbreviations. IPLC, Initial primary lung cancer; SPLC, second primary lung cancer; MEC, the Multiethnic Cohort Study; AUC, Area Under the Curve.



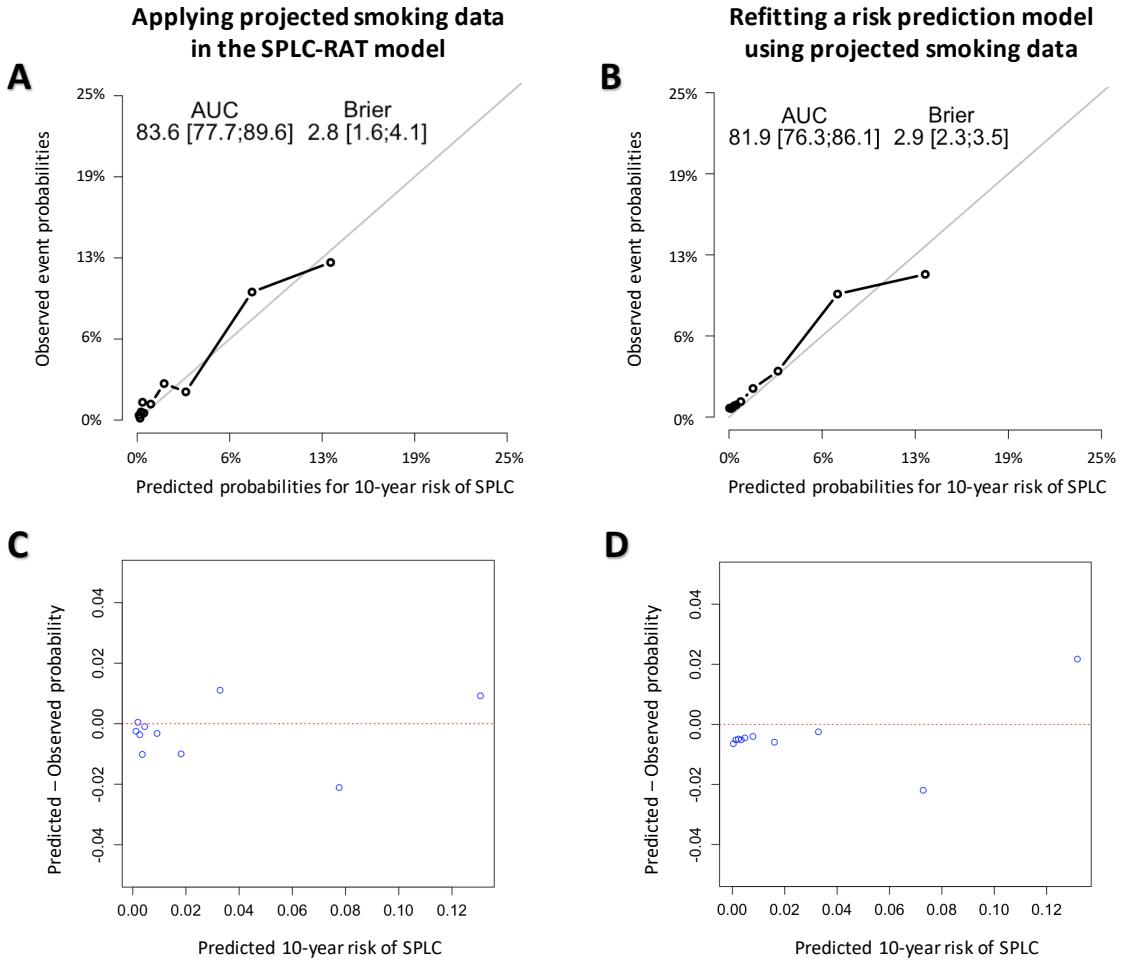
Supplementary Figure 5. Sensitivity analysis for evaluating the performance of the re-estimated SPLC model using a Fine-Gray subdistribution hazards model in the MEC.
The parameters of the proposed SPLC prediction model were re-estimated using a Fine-Gray subdistribution hazard regression (FGR) (see **Supplementary Table 4**) that was used for performance evaluation. (A) Calibration plot (Black) estimated from FGR, together with calibration plot (Red) from the proposed SPLC prediction model, built based on Cause-Specific Cox proportional hazards regression (CSC) for comparison (see **Table 2** and **Figure 1**). (B) Calibration error of mean difference between observed and predicted 10-year SPLC probabilities by risk deciles of the FGR model. All estimates of predictive accuracy were bias corrected based on 200 bootstrap validation. Abbreviations. AUC, Area Under the Curve; FGR, Fine-Gray subdistribution hazard regression; CSC, Cause-specific Cox hazard regression.



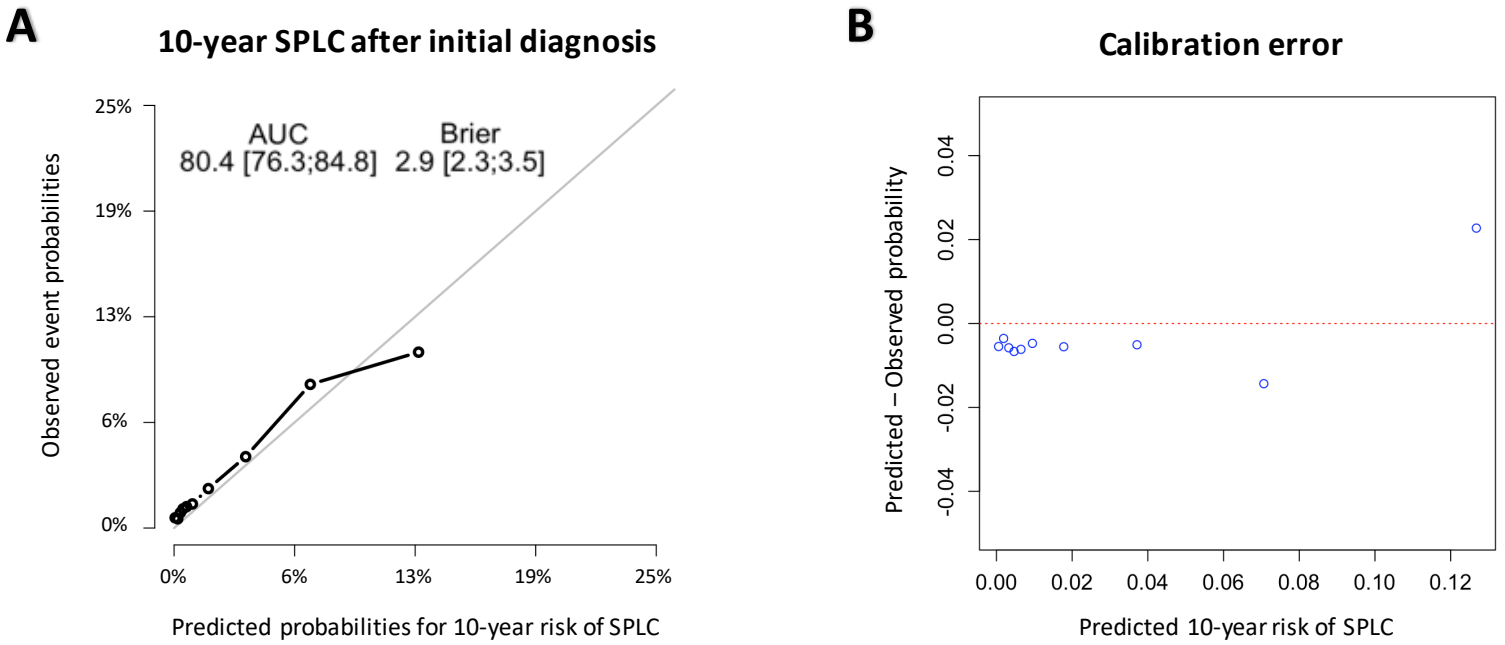
Supplementary Figure 6. Sensitivity analysis for evaluating the performance of the re-estimated SPLC model using multiple imputation in the MEC.
The parameters of the proposed SPLC prediction model were re-estimated using multiple imputation (see **Supplementary Table 5**) that was used for performance evaluation. (A) Calibration plot with discriminative performance (area under the curve; AUC) and prediction accuracy (Brier score). (B) Calibration error of mean difference between observed and predicted 10-year SPLC probabilities by risk deciles. All estimates of predictive accuracy were bias corrected based on 200 bootstrap validation. Abbreviations. AUC, Area Under the Curve; SPLC, second primary lung cancer; MEC, the Multiethnic Cohort Study.



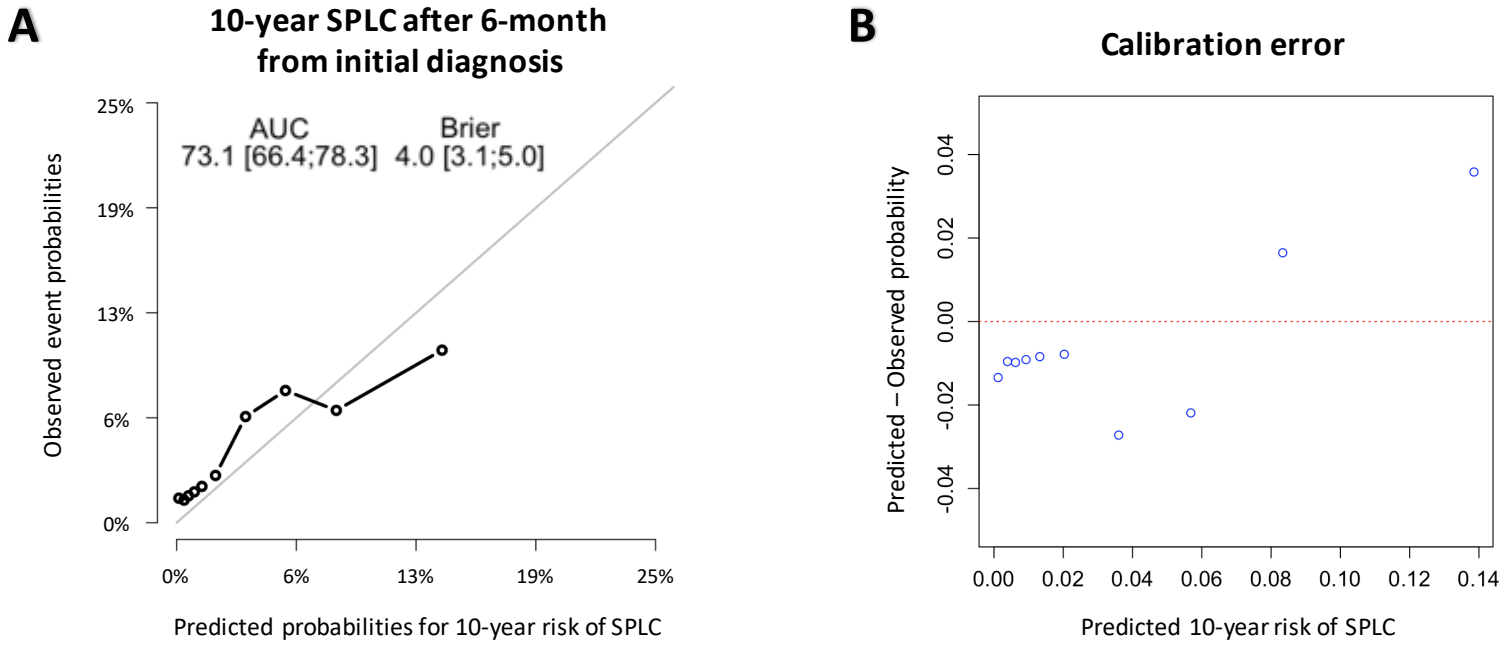
Supplementary Figure 7. Sensitivity analysis for evaluating the performance of the proposed and re-estimated SPLC models using the projected smoking data in the MEC.
(A) Calibration plots and (C) calibration errors with discriminative performance (area under the curve; AUC) and Brier score of the proposed SPLC model (see **Table 2**) by applying the projected smoking data instead of the observed smoking data that was used for model development. (B) Calibration plot and (D) calibration error with AUC and Brier score of the re-estimated SPLC prediction model (see **Supplementary Table 6**) based on the projected smoking data; the estimates of predictive accuracy were bias corrected based on 200 bootstrap validation. Abbreviations. AUC, Area Under the Curve; SPLC, second primary lung cancer; MEC, the Multiethnic Cohort Study



Supplementary Figure 8. Sensitivity analysis for evaluating the performance of the proposed SPLC model using data from all IPLC patients that include never-smoking IPLC cases in the MEC (5,354 ever-smoking IPLC cases and 740 never-smoking IPLC cases). (A) Calibration plots with discriminative performance (area under the curve; AUC) and prediction accuracy (Brier score) of the proposed SPLC prediction model (see **Table 2**). (B) Calibration error of mean difference between observed and predicted 10-year SPLC probabilities by risk deciles in Panel B. All estimates of predictive accuracy were bias corrected based on 200 bootstrap validation. Abbreviations. AUC, Area Under the Curve; IPLC, initial primary lung cancer; SPLC, second primary lung cancer; MEC, the Multiethnic Cohort Study.



Supplementary Figure 9. Sensitivity analysis for evaluating the performance of the re-estimated SPLC model using data excluding 2,098 patients who died or were lost to follow-up within 6 months after IPLC diagnosis in the MEC.
The parameters of the proposed SPLC prediction model were re-estimated using data excluding 2,098 patients who died or were lost to follow-up within 6 months after IPLC diagnosis in the MEC (see **Supplementary Table 8**). (A) Calibration plot with discriminative performance (area under the curve; AUC) and prediction accuracy (Brier score). (B) Calibration error of mean difference between observed and predicted 10-year SPLC probabilities by risk deciles. All estimates of predictive accuracy were bias corrected based on 200 bootstrap validation. Abbreviations. AUC, Area Under the Curve; IPLC, initial primary lung cancer; SPLC, second primary lung cancer; MEC, the Multiethnic Cohort Study.



Supplementary Figure 10. R Shiny application for the SPLC-RAT model.
The model is open for free public use and can be accessed at: https://splc-risk-prediction.shinyapps.io/SPLC-RiskAssessmentTool/