

Risk Stratification for Second Primary Lung Cancer

Summer S. Han, Gabriel A. Rivera, Martin C. Tammemägi, Sylvia K. Plevritis, Scarlett L. Gomez, Iona Cheng, and Heather A. Wakelee

Author affiliations and support information (if applicable) appear at the end of this article.

Published at jco.org on June 23, 2017.

Corresponding author: Heather A. Wakelee, MD, Stanford University, 875 Blake Wilbur Dr, Room 2233, Stanford, CA 94305-5826; e-mail: hwakelee@stanford.edu.

© 2017 by American Society of Clinical Oncology

0732-183X/17/3525w-2893w/\$20.00

A B S T R A C T

Purpose

This study estimated the 10-year risk of developing second primary lung cancer (SPLC) among survivors of initial primary lung cancer (IPLC) and evaluated the clinical utility of the risk prediction model for selecting eligibility criteria for screening.

Methods

SEER data were used to identify a population-based cohort of 20,032 participants diagnosed with IPLC between 1988 and 2003 and who survived ≥ 5 years after the initial diagnosis. We used a proportional subdistribution hazards model to estimate the 10-year risk of developing SPLC among survivors of lung cancer LC in the presence of competing risks. Considered predictors included age, sex, race, treatment, histology, stage, and extent of disease. We examined the risk-stratification ability of the prediction model and performed decision curve analysis to evaluate the clinical utility of the model by calculating its net benefit in varied risk thresholds for screening.

Results

Although the median 10-year risk of SPLC among survivors of LC was 8.36%, the estimated risk varied substantially (range, 0.56% to 14.3%) when stratified by age, histology, and extent of IPLC in the final prediction model. The stratification by deciles of estimated risk showed that the observed incidence of SPLC was significantly higher in the tenth-decile group (12.5%) versus the first-decile group (2.9%; $P < 10^{-10}$). The decision curve analysis yielded a range of risk thresholds (1% to 11.5%) at which the clinical net benefit of the risk model was larger than those in hypothetical all-screening or no-screening scenarios.

Conclusion

The risk stratification approach in SPLC can be potentially useful for identifying survivors of LC to be screened by computed tomography. More comprehensive environmental and genetic data may help enhance the predictability and stratification ability of the risk model for SPLC.

J Clin Oncol 35:2893-2899. © 2017 by American Society of Clinical Oncology

INTRODUCTION

Lung cancer (LC) is the leading cause of cancer-related mortality in the United States.¹ The National Lung Screening Trial recently showed that low-dose computed tomography (LDCT) screening is effective in reducing LC mortality by 20% compared with chest x-ray screening.² With the adoption of computed tomography screening, the number of survivors of LC is expected to rapidly increase.³ The estimated number of survivors of LC was approximately 412,230 as of 2012 and is projected to be $> 500,000$ in 2022.³ Recent studies have shown that survivors of LC have a high risk of developing second primary lung cancer (SPLC), with an incidence four- to six-times higher than that of initial primary lung cancer (IPLC).⁴

Despite the increasing importance of SPLC, there is uncertainty on how to guide screenings for survivors of LC. Recently, the US Preventive Services Task Force established national lung screening guidelines for IPLC, recommending that asymptomatic persons 55 to 80 years of age with ≥ 30 pack-years of smoking and < 15 years since smoking cessation be screened annually by LDCT.⁵

However, there currently are no consensus screening guidelines for survivors of LC who are at a high risk of SPLC. Although several recent studies showed that risk model-based screening is important to efficiently detect IPLC,⁶⁻⁸ such approaches have never been examined for SPLC. To implement effective screening programs for survivors of LC, it is essential to identify factors associated with SPLC risk and to evaluate

ASSOCIATED CONTENT



Appendix
DOI: <https://doi.org/10.1200/JCO.2017.72.4203>

DOI: <https://doi.org/10.1200/JCO.2017.72.4203>

individuals' risk factors for SPLC that can help identify efficient screening criteria.

Although numerous studies have examined the risk of SPLC, many of them focused on estimating the cumulative incidence of SPLC in study populations,⁹⁻¹² and the factors that contribute to SPLC risk have not been established. Johnson⁹ analyzed 10 published studies to examine the cumulative risk of SPLC; this work showed that nine of the 10 studies reported a rate of SPLC risk of 1% to 2% per patient per year among patients with resected non-small-cell lung cancer. The factors associated with SPLC risk were not reported, nor were prediction models provided for survivors of IPLC. Whereas a recent study by Boyle et al¹³ reported that smoking is a risk factor for SPLC, another study by Ripley et al¹⁴ reported no association between smoking and SPLC risk. Most published studies for SPLC, including these two studies, report results from the experiences of single institutions¹²⁻¹⁴; these results may be heterogeneous, with a lack of power to identify the important factors associated with SPLC.

In this study, we aimed to evaluate the risk of SPLC among patients with LC who survived ≥ 5 years after the diagnosis of IPLC. We used data from a large population-based cohort to identify the clinical and demographic factors associated with SPLC risk. In a prior study, we compared various statistical modeling approaches to analyze SPLC risk; we showed that application of the standard Cox regression, which does not take into account competing risks, leads to a substantial bias in risk estimation of SPLC (Han et al, Competing risk analysis of SPLC). In an extension of this work, we applied a proportional subdistribution hazards regression¹⁵ to obtain unbiased estimates of SPLC risk in the presence of competing risks. In addition, we evaluated the risk-stratification ability and the clinical utility of the prediction model by using decision curve analysis.

METHODS

SEER Data

We obtained the study participants from the population-based SEER program of the National Cancer Institute. With a focus on evaluation of the risk of developing SPLC among survivors of IPLC, in accordance with clinical opinions, we defined survivors of LC as patients with IPLC who survived ≥ 5 years after the diagnosis of IPLC. We identified 20,032 patients who were diagnosed with IPLC between 1988 and 2003, and who survived ≥ 5 years after the initial diagnosis. We focused on this period because it was before active computed tomography screenings were adopted, and hence most cancers were likely to be detected clinically in the absence of screening; this can help avoid potential bias that could be raised by allowing different detection modes across patients.

Staging and Histology

Histologic classification was according to the International Classification of Diseases of Oncology, 3rd edition. All tumors were staged according to the American Joint Committee on Cancer staging system. Collected clinical variables included stage, histology, tumor size, extent of disease, first-course treatment, survival of up to 16 years after the diagnosis of IPLC, and the incidence of SPLC. Demographic variables such as age at lung cancer diagnosis, sex, and race/ethnicity were also collected; smoking information was not available in SEER data. We only included histologically confirmed malignant tumors of the lung and excluded patients with missing information on stage, race/ethnicity, age, histology, and treatment.

Definition of SPLC

The most widely used definition of SPLC, proposed by Martini and Melamed,¹⁶ considers a new, distinct pulmonary malignancy to be SPLC if it fulfills any one of the following three criterion: (1) histologic results are different from those of IPLC; (2) the histologic results are the same as for the index tumor but diagnosed 2 years after the primary tumor; or (3) the histologic results are the same as for the index tumor, diagnosed within 2 years of the primary tumor, but are located in different lobes or segments, with no positive intervening lymph nodes and no evidence of metastasis. Reflecting the concerns on the complexity in diagnosis of SPLC,^{17,18} however, we considered alternative criteria that are stricter, in accordance with clinical opinions, where SPLC is defined as a new, distinct pulmonary malignancy that is diagnosed ≥ 5 years after the primary tumor.

Statistical Methods

Proportional subdistribution hazards modeling. Although the most common method for analyzing survival data is the Cox proportional hazards regression model, it is not designed to accommodate the competing nature of multiple causes of the same event; hence, it tends to produce inaccurate estimates.¹⁹

Competing risks are especially relevant in the study of SPLC, where a substantial proportion of survivors of LC often die as the result of other causes (such as heart disease) before developing SPLC. To obtain unbiased estimates of the risk of SPLC, we applied a proportional subdistribution hazards regression,¹⁵ which connects regression coefficients to a cumulative incidence function to estimate the unbiased risks in the presence of competing risks.

Variable selection. The following variables were considered for predicting the risk of SPLC: sex, race, age at IPLC diagnosis, stage, histology, disease extent (Table 1), tumor size, node involvement, the number of positive nodes, and first course of treatment. We also considered interactions of all possible pairs of variables. To select variables to be included in the final prediction model, we used stepwise forward and backward elimination methods. Stepwise selection is a method of fitting models in which the selection of predictive variables is performed by an automatic procedure. In each step, a variable is considered for addition to or subtraction from the set of variables on the basis of some prespecified criterion. We used the Akaike information criterion and the Bayesian information criterion (BIC) as well as a new criterion called BICcr, which is an adjustment of BIC to competing risks models.²⁰ We quantified variable importance by conducting likelihood ratio tests and using χ^2 test statistics subtracted by the degrees of freedom as a comparison metric.²¹

Validation and performance evaluation. As a validation study, we used a bootstrap cross-validation method; risk prediction models were trained on 200 bootstrap samples that were drawn with replacement of the same size as the original data. We then evaluated the performance of the models in the observations that were not in the bootstrap sample to estimate unbiased estimate of performance metrics. To evaluate the performance of the prediction model, we used the c-index and calibration plot. The c-index measures discrimination, which is the ability of a model to distinguish subject outcomes. The calibration of a prediction model measures the overall agreement between the observed outcomes and the predicted probability using the model.

Risk stratification ability. Given recent criticisms of the c-index,²²⁻²⁵ we considered another way to evaluate a risk prediction model—the ability to stratify a population into groups with distinct risks that can substantially affect the risk-benefit balance of screening.²⁴ As more predictive risk factors are identified and incorporated into a model, estimated risks will have more variations among individuals, which can be useful for identifying high-risk individuals for disease prevention. To examine the risk stratification ability of the prediction model for SPLC, we divided the study population into 10 groups by deciles on the basis of estimated risk. We then estimated observed cumulative incidence (10 years) using the Gray method²⁶ for each group by taking into account competing risks and comparing them across the deciles.

Clinical utility: Decision curve analysis. Although the c-index is one of the most commonly used methods to evaluate the performance of

Risk Stratification for Second Primary Lung Cancer

Table 1. Patient Characteristics of the Study Population in SEER

Variable	Stratified Events, No. (%)			
	Censored	SPLC	Death From IPLC	Death From Other Causes
Total No. of events	11,680 (58.3)	1,018 (5.1)	2,554 (12.7)	4,780 (23.9)
Sex				
Female	6,264 (53.6)	487 (47.8)	1,302 (51.0)	2,205 (46.1)
Male	5,416 (46.4)	531 (52.2)	1,252 (49.0)	2,575 (53.9)
Race				
Asian Pacific	741 (6.3)	55 (5.4)	140 (5.5)	202 (4.2)
Black	1,024 (8.8)	102 (10.0)	249 (9.7)	383 (8.0)
Other	51 (0.4)	2 (0.2)	7 (0.3)	13 (0.3)
White	9,864 (84.5)	859 (84.4)	2,158 (84.5)	4,182 (87.5)
Histology of IPLC				
AD	3,832 (32.8)	343 (33.7)	894 (35.0)	1,420 (29.7)
SQ	2,471 (21.2)	279 (27.4)	618 (24.2)	1,458 (30.5)
LC	652 (5.6)	75 (7.4)	169 (6.6)	294 (6.2)
SC	627 (5.4)	77 (7.6)	203 (7.9)	309 (6.5)
Other	4,098 (35.1)	244 (24.0)	670 (26.2)	1,299 (27.2)
Age at diagnosis of IPLC, years				
0-44	2 (7.7)	26 (2.6)	74 (2.9)	64 (1.3)
45-49	807 (6.9)	51 (5.0)	87 (3.4)	103 (2.2)
50-54	1,273 (10.9)	84 (8.3)	187 (7.3)	213 (4.5)
55-59	1,631 (14.0)	147 (14.4)	325 (12.7)	399 (8.3)
60-64	1,990 (17.0)	238 (23.4)	450 (17.6)	691 (14.5)
65-69	2,181 (18.7)	241 (23.7)	563 (22.0)	1,045 (21.9)
70-74	1,681 (14.4)	163 (16.0)	473 (18.5)	1,135 (23.7)
75-79	880 (7.5)	53 (5.2)	268 (10.5)	743 (15.5)
80-84	281 (2.4)	14 (1.4)	101 (4.0)	312 (6.5)
≥ 85	54 (0.5)	1 (0.1)	26 (1.0)	75 (1.6)
Node involvement of IPLC				
No involvement	9,031 (77.3)	791 (77.7)	1,672 (65.5)	3,593 (75.2)
Regional	2,583 (22.1)	221 (21.7)	862 (33.8)	1,158 (24.2)
Distant	66 (0.6)	6 (0.6)	20 (0.8)	29 (0.6)
Extent of IPLC (%)*				
Localized	8,143 (69.7)	743 (73.0)	1,574 (61.6)	3,347 (70.0)
Regional	3,132 (26.8)	252 (24.8)	823 (32.2)	1,258 (26.3)
Distant	405 (3.5)	23 (2.3)	157 (6.1)	175 (3.7)
IPLC tumor size, cm				
< 2	4,172 (35.7)	364 (35.8)	758 (29.7)	1,555 (32.5)
2-3	1,811 (15.5)	149 (14.6)	347 (13.6)	716 (15.0)
3-4	3,086 (26.4)	293 (28.8)	719 (28.2)	1,392 (29.1)
> 4	2,611 (22.4)	212 (20.8)	730 (28.6)	1,117 (23.4)
Positive nodes of IPLC				
0	9,132 (78.2)	796 (78.2)	1,708 (66.9)	3,621 (75.8)
1	1,635 (14.0)	151 (14.8)	537 (21.0)	737 (15.4)
2	483 (4.1)	37 (3.6)	152 (6.0)	219 (4.6)
> 3	430 (3.7)	34 (3.3)	157 (6.1)	203 (4.2)
Radiation therapy of IPLC				
No radiation	9,274 (79.4)	808 (79.4)	1,619 (63.4)	3,635 (76.0)
Radiation	2,406 (20.6)	210 (20.6)	935 (36.6)	1,145 (24.0)
Stage of IPLC				
Early/I, II	9,298 (79.6)	809 (79.5)	1,716 (67.2)	3,757 (78.6)
Advanced/III, IV	2,382 (20.4)	209 (20.5)	838 (32.8)	1,023 (21.4)
Treatment of IPLC				
No radiation and/or surgery	10,358 (88.7)	905 (88.9)	2,153 (84.3)	4,224 (88.4)
Radiation and surgery	1,322 (11.3)	113 (11.1)	401 (15.7)	556 (11.6)
Mean length of follow-up after IPLC diagnosis, years, ± SD	9.45 ± 2.85	8.49 ± 2.30	7.72 ± 2.00	8.43 ± 2.26

Abbreviations: AD, adenocarcinoma; IPLC, initial primary lung cancer; LC, large cell; SC, small-cell lung cancer; SD, standard deviation; SPLC, second primary lung cancer; SQ, squamous cell.

*Disease extent was defined as follows, according to SEER Extent of Disease – 1998, 3rd edition. Localized: an invasive neoplasm confined entirely to the organ of origin (lung). Regional: a neoplasm that has extended (1) beyond the limits of the organ of origin (lung) directly into surrounding organs or tissues; (2) into regional lymph nodes by way of the lymphatic system; or (3) by a combination of extension and regional lymph nodes. Distant: a neoplasm that has spread to parts of the body remote from the primary tumor either by direct extension or by discontinuous metastasis (eg, implantation or seeding) to distant organs, tissues, or via the lymphatic system to distant lymph nodes.

a prediction model, it does not have a direct clinical interpretation. To address this issue, we applied decision curve analysis, which evaluates the clinical usefulness of a prediction model by calculating its net benefit using

the rate of true and false positives in varied risk thresholds for screening.²⁵ A model is clinically useful in the context of screening if the application of the model produces a larger net benefit than not applying it in identifying

screening eligibility. More details on this method are provided in the Appendix (online only). All statistical analyses were conducted using R (<https://www.r-project.org/>).

RESULTS

Development of a Risk Model: Identifying Factors Associated With SPLC Risk

Characteristics of the study population are listed in Table 1; the maximum follow-up was 16 years and the median follow-up was 8 years from the diagnosis of IPLC. Results for the selected variables are listed in Table 2, which indicate that the age of IPLC diagnosis, histology, and extent of disease were significantly associated with SPLC risk. Figure 1 shows the relative importance of each independent variable, and shows that patient age at the time of IPLC diagnosis was the most important factor in predicting SPLC risk, followed by histology and extent of disease. Compared with a reference age group of 70 to 74 years, survivors of LC with a younger age (< 45 years) or an older age (> 75 years) had substantially reduced risks of SPLC, with a subdistribution hazard ratio (HR_s) < 0.59. Survivors of LC in the “other” histology group had significantly reduced risks versus adenocarcinoma ($HR_s = 0.79$; $P = 4.3 \times 10^{-3}$). Refitting the proportional subdistribution hazards

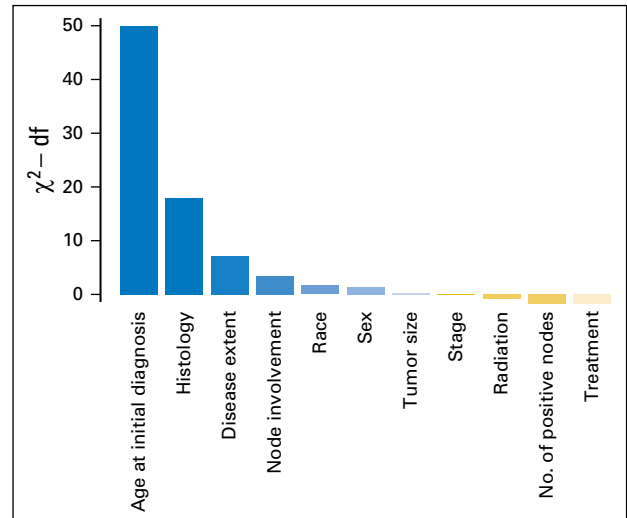


Fig 1. Importance of variables in predicting the risk of second primary lung cancer. The χ^2 axis shows the likelihood ratio test χ^2 statistic subtracted by the degrees of freedom (df) conducted for each variable. Note that none of interaction terms were significant; hence, the importance metric was displayed for the main effects of the variables in the multivariable model.

Table 2. Factors Associated With Second Primary Lung Cancer (SPLC) Risk Among Survivors of Lung Cancer Included in the Final Prediction Model for SPLC Risk

Factor	HR_s	95% CI	<i>P</i>
Histology of IPLC			
AD (reference)	1.00		
SQ	1.06	0.91 to 1.24	.4600
LC	1.15	0.90 to 1.48	.2600
SC	1.22	0.96 to 1.56	.1100
Other*	0.79	0.67 to 0.93	< .001
Age at diagnosis of IPLC, years			
70-75 (reference)	1.00		
0-44	0.54	0.36 to 0.83	< .001
45-49	1.04	0.76 to 1.42	.8000
50-54	1.00	0.77 to 1.30	1.0000
55-59	1.25	1.00 to 1.56	.0520
60-64	1.45	1.19 to 1.76	< .001
65-69	1.24	1.02 to 1.51	< .001
75-79	0.59	0.43 to 0.81	< .001
80-84	0.44	0.26 to 0.76	< .001
≥ 85	0.15	0.02 to 1.06	.0570
Extent of IPLC†			
Localized (reference)	1.00		
Regional	0.85	0.74 to 0.98	< .001
Distant	0.54	0.36 to 0.82	< .001

Abbreviations: AD, adenocarcinoma; HR_s , subdistribution hazard ratio; IPLC, initial primary lung cancer; LC, large cell; SC, small-cell lung cancer; SQ, squamous cell.

*Other includes bronchioloalveolar carcinoma (30%), carcinoid (22%), carcinoma not otherwise defined (18%), adenosquamous (7%), and others (< 5%).

†Disease extent was defined as follows, according to SEER Extent of Disease—1988, 3rd edition. Localized: an invasive neoplasm confined entirely to the organ of origin (lung). Regional: a neoplasm that has extended (1) beyond the limits of the organ of origin (lung) directly into surrounding organs or tissues; (2) into regional lymph nodes by way of the lymphatic system; or (3) by a combination of extension and regional lymph nodes. Distant: a neoplasm that has spread to parts of the body remote from the primary tumor either by direct extension or by discontinuous metastasis (eg, implantation or seeding) to distant organs, tissues, or via the lymphatic system to distant lymph nodes.

model with a more detailed histology variable (nine levels instead of five levels) revealed that carcinoid and carcinoma not otherwise specified in the “other” histology group drove the reduced risks ($HR_s = 0.28$ and 0.68 , respectively; $P = 2.3 \times 10^{-8}$ and $.013$, respectively; Appendix Tables A1 and A2, online only). Regional and distant extensions of IPLC were also associated with a decreased risk of SPLC, potentially as the result of a higher chance of death from IPLC before developing SPLC. Validation of the risk model showed that the c-index for validation was 61.2%, with moderate discriminatory power; in addition, calibration results demonstrated good concordance between the observed incidences and predictions, with the curves following the 45° line (Appendix Fig A1, online only).

Risk Stratification: Variation of SPLC Risk Across Individuals

Using the risk prediction model in Table 2, we estimated an individual's risk of developing SPLC¹⁵ within 10 years after surviving 5 years from their diagnosis of IPLC (ie, 15 years since the diagnosis of IPLC). Although the median 10-year risk for SPLC among survivors of LC was 8.35%, the risk varied substantially (range, 0.59% to 14.3%) when stratified by age, histology, and extent of disease using the risk model. Figures 2A and 2B show that the observed risks varied substantially by age or histology (eg, age group ≥ 85 years had a median 10-year risk of 1.1%, whereas the age group 60 to 64 years had the highest median risk at 10.97% [$P < 10^{-10}$]). The cumulative risk stratified by each percentile of the risk estimated using all selected features, eg, age, histology, and disease extent (Fig 2C), shows a wide stratification of risks, from 1.3% for the bottom first percentile to 12.56% for the top first percentile, demonstrating a good separation among low-risk versus high-risk individuals. We compared the observed incidence of SPLC across different risk groups defined by the deciles of estimated risk in Figure 3, which shows that a significantly higher incidence was observed in the tenth-decile group (12.5%) versus

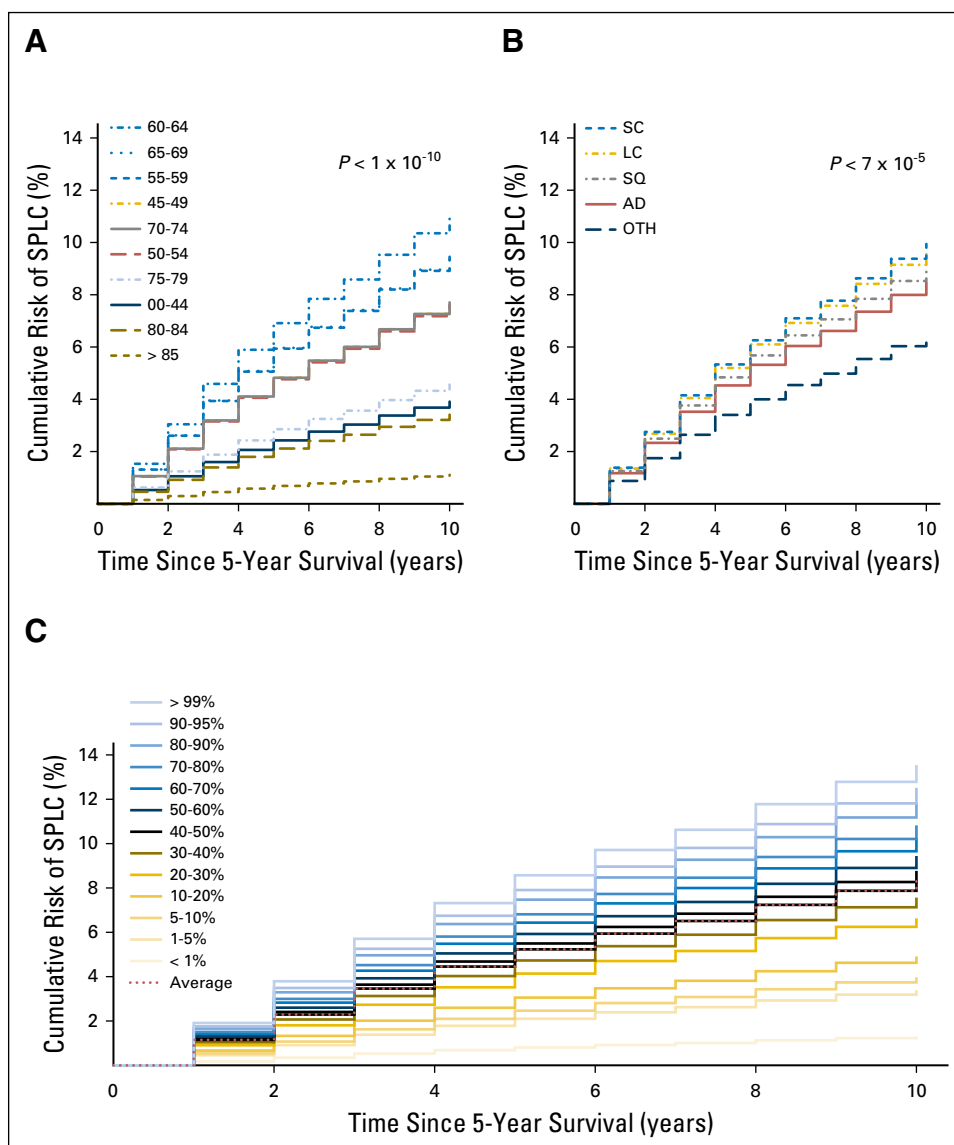


Fig 2. Cumulative risk of second primary lung cancer (SPLC) among survivors of lung cancer (ie, patients with initial primary lung cancer [IPLC] who survived ≥ 5 years after the diagnosis of IPLC). The cumulative risk of SPLC is shown by (A) age group, and by (B) histologic subtype. The risks across different groups were compared and tested using the method by Gray.²⁶ (C) The cumulative risk is shown by percentile of estimated risk on the basis of the prediction model. AD, adenocarcinoma; LC, large cell; OTH, other; SC, small-cell lung cancer; SQ, squamous cell.

the first-decile group (2.9%; $P < 10^{-10}$). This suggests that the risk stratification for SPLC using the risk model can be potentially useful in identifying high-risk individuals among survivors of LC.

Evaluating the Clinical Utility of the Risk Prediction Model

We compared the net benefit of the risk model to those in two alternative scenarios: screening all individuals and screening no one. The results shown in Figure 4 demonstrate that in a wide range of threshold probabilities (1% to 11%), the clinical net benefit of the risk model was larger than that in hypothetical all-screening or no-screening scenarios. This implies that if we use a risk threshold from the given interval of 1% to 11% (eg, 5%), so that screening is recommended if an individual's risk is above the given threshold, then the calculated net benefit (the weighted sum of true positives subtracted by the number of false positives) is larger for the prediction model than it is in the strategies that do not use the model.

DISCUSSION

Using population-based cancer surveillance data, we evaluated an individual's risk of developing SPLC among patients who survived ≥ 5 years after the diagnosis of IPLC. Although the median 10-year risk for SPLC was 8.35% for the entire sample, the estimated risk varied substantially across individuals (range, 0.59% to 14.3%) when stratified by various risk factors. The stratification of patients by deciles of estimated risk showed that the observed incidence of SPLC was significantly higher in the tenth-decile group compared with the first-decile group, implying that the risk model provides a good separation of high- versus low-risk survivors. Unlike most prior studies that have assessed overall risk of SPLC among patients who have undergone resection,^{9,12,14} our study focused on individuals who survived ≥ 5 years after the diagnosis of IPLC. Although patients are under intense surveillance for potential recurrence or metastases after resection, those who survive long enough (eg, ≥ 5 years) are often in high need of

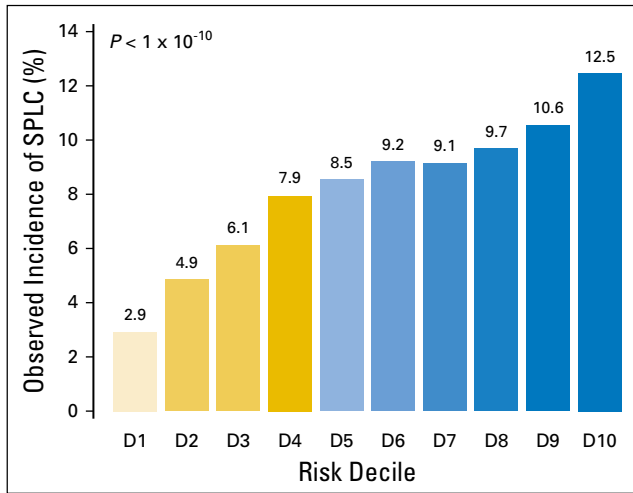


Fig 3. Cumulative 10-year incidence of second primary lung cancer (SPLC) by decile (D) of the estimated risk using the prediction model. In each decile of individuals, the marginal cumulative incidence of SPLC was calculated and the equality of the estimated incidences across groups was tested using the method by Gray.²⁶

guidance on screening and can benefit from informed decision making.

To the best of our knowledge, this is the first study to evaluate the risk stratification ability and clinical utility of a prediction model for SPLC among survivors of LC using data from a large, population-based cohort. Unlike studies conducted using data from single-institution based studies, this allows the estimation of a number of risk factors on the basis of a large sample that is not subject to selection and referral biases. It is notable that all SEER registries comply with the strictest data-quality indicators and follow consistent criteria for collecting data and maintaining the quality of its variables. Recently, our group evaluated and compared the

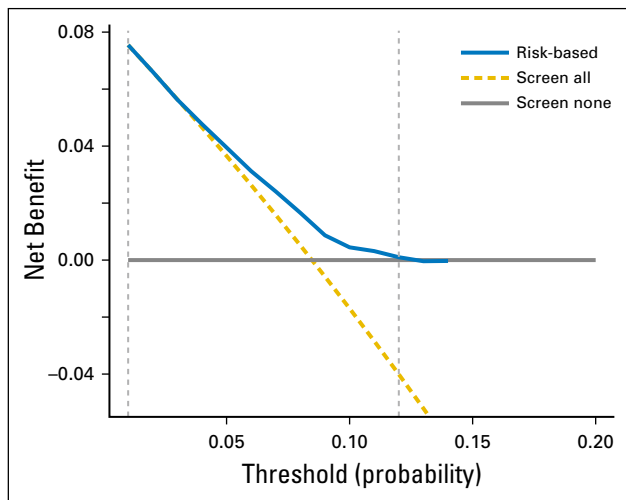


Fig 4. Decision curve analysis for the risk model for second primary lung cancer. The x-axis is the risk threshold probability that changes from 0 to 1 (right truncated at 0.2) and the y-axis is the calculated net benefit for a given threshold probability. The blue curve depicts the net benefit of the risk model-based selection strategy for screening, whereas the gold and gray lines display the net benefits in the alternative strategies of screening all patients (gold) versus screening no patients (gray) in the data set.

performance of various statistical modeling approaches to estimate the risk of SPLC (Han et al, Competing risk analysis of SPLC). However, the risk stratification ability of the prediction model has not been investigated nor was the examination of the clinical utility of the risk model reported. In the current study, we quantified the clinical net benefits of the risk model-based screening strategy for SPLC.

Research on the optimal risk threshold for screening for IPLC is ongoing. Recently, Tammemägi et al⁸ proposed a 6-year risk threshold of 1.5% for IPLC risk for screening, which represents the 65th percentile of the 6-year risk among the high-risk individuals who participated in the National Lung Screening Trial. A direct comparison may not be appropriate because of the application of different risk models. However, the 65th percentile of the 6-year risk for SPLC that we estimated for survivors of LC in SEER data, using our model, was 6.2%, which is more than four times higher than that of IPLC. This finding is consistent with prior reports that compared the incidences of SPLC versus IPLC.⁴ Interestingly, using the 1.5% threshold recommended for IPLC, 99% of the survivors of LC in the SEER data we used were eligible for screening (data not shown). This implies that a tailored prediction model for SPLC is needed for survivors of LC to stratify a population into groups with distinct risks to substantially affect the risk-benefit balance of screening.

The association between young age and lower risk of several second cancers has been reported previously.²⁷ Although this finding is rather counterintuitive because young survivors have more time to develop SPLC and thus could have a higher risk of SPLC, it is possible that this association exists partly because young patients with LC have different smoking behaviors. Although SEER data do not provide information on smoking, we compared the histology distribution among young (age < 45 years) versus middle-aged (55 to 75 years) patients and found that the proportion of squamous LC (known to be associated with heavy smoking) was much lower among younger patients (9% v 27%, respectively). It is possible that the etiology of LC among a young population may be different from that of smoking-related LC, with potentially different driver mutations that may be more curable or that may lower the chance of developing SPLC. Further investigation into this hypothesis is necessary, but will require additional databases that contain information on driver mutations.

Despite several strengths, our study has limitations. First, SEER does not provide environmental exposures data, including smoking. Although smoking is an established risk factor for IPLC, its effect on SPLC is not clear. Given that a high proportion of patients with LC are already smokers, the effect of smoking on SPLC among patients with LC may not be as strong as it is on the risk of IPLC. Recently, Ripley et al¹⁴ showed that smoking is not associated with SPLC, but this finding needs to be confirmed in a larger study.

SEER data also do not include other potentially important information for SPLC, such as family history of LC, chronic obstructive pulmonary disease, or genetic variants. It is possible that the lack of these data has led to the moderate c-index observed in our model. However, although the c-index is currently considered to be the standard metric to evaluate the accuracy of risk prediction models, it is criticized for having no direct clinical relevance and

for avoiding the supposed subjectivity in the threshold selection by summarizing overall model performance above all possible thresholds.²²⁻²⁵ Despite the moderate *c*-index, our model has shown the ability to stratify a population into subpopulations with distinct risks that can potentially effect the risk-benefit balance of screening. It has also proven to be clinically useful in providing screening eligibility criteria compared with strategies that do not use models based on decision curve analysis.

In conclusion, we developed a prediction model for SPLC risk on the basis of clinical and demographic risk factors using data from a large population-based cohort. Our analysis shows that the risk stratification approach for SPLC can be potentially useful in identifying survivors of LC at high risk who should be screened using LDCT. Future directions include validating and extending the proposed model using more comprehensive environmental and genetic data, which can help enhance the predictability and stratification ability of the risk model for SPLC.

REFERENCES

1. Siegel R, Ma J, Zou Z, et al: Cancer statistics, 2014. *CA Cancer J Clin* 64:9-29, 2014
2. Aberle DR, Adams AM, Berg CD, et al: Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 365:395-409, 2011
3. Siegel R, DeSantis C, Virgo K, et al: Cancer treatment and survivorship statistics, 2012. *CA Cancer J Clin* 62:220-241, 2012
4. Surapaneni R, Singh P, Rajagopalan K, et al: Stage I lung cancer survivorship: Risk of second malignancies and need for individualized care plan. *J Thorac Oncol* 7:1252-1256, 2012
5. Moyer VA; U.S. Preventive Services Task Force: Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* 160:330-338, 2014
6. Tammemägi MC, Katki HA, Hocking WG, et al: Selection criteria for lung-cancer screening. *N Engl J Med* 368:728-736, 2013
7. Kovalchik SA, Tammemägi M, Berg CD, et al: Targeting of low-dose CT screening according to the risk of lung-cancer death. *N Engl J Med* 369:245-254, 2013
8. 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* 11:e1001764, 2014
9. Johnson BE: Second lung cancers in patients after treatment for an initial lung cancer. *J Natl Cancer Inst* 90:1335-1345, 1998

10. Rice D, Kim H-W, Sabichi A, et al: The risk of second primary tumors after resection of stage I nonsmall cell lung cancer. *Ann Thorac Surg* 76:1001-1007, discussion 1007-1008, 2003
11. Heyne KH, Lippman SM, Lee JJ, et al: The incidence of second primary tumors in long-term survivors of small-cell lung cancer. *J Clin Oncol* 10:1519-1524, 1992
12. 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* 145:75-81, discussion 81-82, 2013
13. Boyle JM, Tandberg DJ, Chino JP, et al: Smoking history predicts for increased risk of second primary lung cancer: A comprehensive analysis. *Cancer* 121:598-604, 2015
14. Ripley RT, McMillan RR, Sima CS, et al: Second primary lung cancers: Smokers versus non-smokers after resection of stage I lung adenocarcinoma. *Ann Thorac Surg* 98:968-974, 2014
15. Fine JP, Gray RJ: A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc* 94:496-509, 1999
16. Martini N, Melamed MR: Multiple primary lung cancers. *J Thorac Cardiovasc Surg* 70:606-612, 1975
17. van Bodegom PC, Wagenaar SS, Corrin B, et al: Second primary lung cancer: Importance of long term follow up. *Thorax* 44:788-793, 1989
18. Xue X, Liu Y, Pan L, et al: Diagnosis of multiple primary lung cancer: A systematic review. *J Int Med Res* 41:1779-1787, 2013

19. Scrucca L, Santucci A, Aversa F: Regression modeling of competing risk using R: An in depth guide for clinicians. *Bone Marrow Transplant* 45:1388-1395, 2010
20. Kuk D, Varadhan R: Model selection in competing risks regression. *Stat Med* 32:3077-3088, 2013
21. Harrell FE Jr. *Regression Modeling Strategies*. Springer, New York, NY, 2015
22. Cook NR, Ridker PM: Advances in measuring the effect of individual predictors of cardiovascular risk: The role of reclassification measures. *Ann Intern Med* 150:795-802, 2009
23. Lobo JM, Jiménez-Valverde A, Real R: AUC: A misleading measure of the performance of predictive distribution models. *Glob Ecol Biogeogr* 17:145-151, 2008
24. Chatterjee N, Shi J, García-Closas M: Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat Rev Genet* 17:392-406, 2016
25. Vickers AJ, Elkin EB: Decision curve analysis: A novel method for evaluating prediction models. *Med Decis Making* 26:565-574, 2006
26. Gray RJ: A class of K-sample tests for comparing the cumulative incidence of a competing risk. *Ann Stat* 16:1141-1154, 1988
27. Park SM, Lim MK, Jung KW, et al: Prediagnosis smoking, obesity, insulin resistance, and second primary cancer risk in male cancer survivors: National Health Insurance Corporation Study. *J Clin Oncol* 25:4835-4843, 2007

Affiliations

Summer S. Han, Sylvia K. Plevritis, and Heather A. Wakelee, Stanford University School of Medicine; Summer S. Han, Sylvia K. Plevritis, Scarlett L. Gomez, Iona Cheng, and Heather A. Wakelee, Stanford Cancer Institute; Heather A. Wakelee and Gabriel A. Rivera, Stanford University Department of Medicine, Division of Oncology, Stanford; Gabriel A. Rivera, Kaiser Permanente Fresno Medical Center, Fresno; Scarlett L. Gomez and Iona Cheng, Cancer Prevention Institute of California, Fremont, CA; and Martin C. Tammemägi, Brock University, St Catharines, Ontario, Canada

Support

Supported by NIH grants U01-CA152956 and U01-CA199284.

Prior Presentation

Presented at the International Association for the Study of Lung Cancer Chicago Multidisciplinary Symposium in Thoracic Oncology, Chicago, IL, September 22-24, 2016.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Disclosures provided by the authors are available with this article at jco.org.

AUTHOR CONTRIBUTIONS

Conception and design: Summer S. Han, Gabriel A. Rivera, Sylvia K. Plevritis, Heather A. Wakelee

Financial support: Sylvia K. Plevritis, Heather A. Wakelee

Collection and assembly of data: Summer S. Han, Gabriel A. Rivera, Heather A. Wakelee

Data analysis and interpretation: Summer S. Han, Martin C.

Tammemägi, Scarlett L. Gomez, Iona Cheng, Heather A. Wakelee

Manuscript writing: All authors

Final approval of manuscript: All authors

Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST**Risk Stratification for Second Primary Lung Cancer**

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/jco/site/ife.

Summer S. Han

No relationship to disclose

Gabriel A. Rivera

No relationship to disclose

Martin C. Tammemägi

No relationship to disclose

Sylvia K. Plevritis

Consulting or Advisory Role: GRAIL

Scarlett L. Gomez

Employment: Abgenix (I), BioInspire (I), Boehringer Ingelheim (I)

Stock or Other Ownership: Amgen (I)

Research Funding: Genentech (Inst)

Travel, Accommodations, Expenses: Genentech (through employment [CPIC]) (I)

Iona Cheng

No relationship to disclose

Heather A. Wakelee

Honoraria: Peregrine Pharmaceuticals, Biocon, ACEA Biosciences, Pfizer, Novartis

Consulting or Advisory Role: Peregrine Pharmaceuticals, ACEA Biosciences, Pfizer, Novartis

Research Funding: Genentech (Inst), Pfizer (Inst), Eli Lilly (Inst), Celgene (Inst), AstraZeneca/MedImmune (Inst), Exelixis (Inst), Novartis (Inst), Regeneron (Inst), Clovis Oncology (Inst), Xcovery (Inst), Bristol-Myers Squibb (Inst), Agennix (Inst), Gilead Sciences (Inst), Pharmacyclics (Inst)

Travel, Accommodations, Expenses: Clovis Oncology, Novartis, ACEA Biosciences, Pfizer, Genentech

Appendix

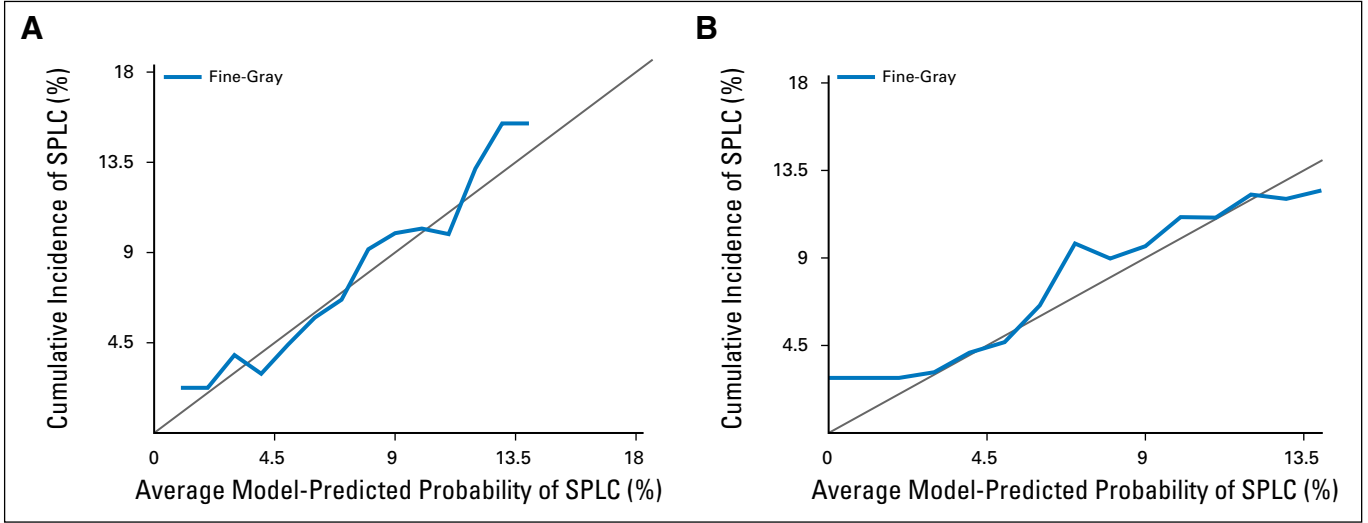


Fig A1. Calibration plots. (A) The calibration on the entire data used to fit the model. (B) The validation of the prediction model using bootstrap cross-validation method. The x-axis shows the mean predicted probability of the conditional cumulative incidence model. The y-axis indicates marginal cumulative incidence probabilities for the respective cohorts. The gray line represents equality between the predicted and observed marginal cumulative incidences.

Histology of IPLC	No.	Relative Frequency
AD	6,489	0.32
SQ	4,826	0.24
SC	1,216	0.06
LC	1,190	0.06
BAC	1,888	0.09
Carcinoid	1,550	0.08
Carcinoma, NOS	1,133	0.06
Adenosquamous	422	0.02
Other	1,318	0.07
Total	20,032	1.00

Abbreviations: AD, adenocarcinoma; BAC, bronchioloalveolar carcinoma; IPLC, initial primary lung cancer; LC, large cell; NOS, not otherwise specified; SC, small-cell lung cancer; SQ, squamous cell.

Table A2. Refitting the Model for Second Primary Lung Cancer With an Extended Histology Variable

Variable	HR _s	95% CI	P
Histology of IPLC			
AD (reference)	1.00		
SQ	1.22	0.80 to 1.76	.1100
LC	1.07	0.91 to 1.40	.4100
SC	1.15	0.18 to 0.43	.2700
BAC	1.13	0.44 to 0.91	.2800
Carcinoid	0.28	0.90 to 1.47	< .001
Carcinoma NOS	0.63	0.63 to 1.10	< .001
Adenosquamous	1.19	0.95 to 1.56	.3900
Other	0.83	0.91 to 1.25	.2000
Age of IPLC diagnosis, years			
70-75 (reference)		1.00	
0-44	0.67	0.44 to 1.03	< .001
45-49	1.13	0.83 to 1.54	.4400
50-54	1.03	0.79 to 1.34	< .001
55-59	1.27	1.02 to 1.59	.0350
60-64	1.45	1.19 to 1.77	.0002
65-69	1.24	1.02 to 1.52	.0300
75-79	0.59	0.43 to 0.80	< .001
80-84	0.44	0.26 to 0.77	< .001
≥ 85	0.15	0.02 to 1.08	.0600
Extent of IPLC*			
Local (reference)	1.00		
Regional	0.83	0.72 to 0.96	.0097
Distant	0.56	0.37 to 0.84	< .001

Abbreviations: AD, adenocarcinoma; BAC, bronchioloalveolar carcinoma; HR_s, subdistribution hazard ratio; IPLC, initial primary lung cancer; LC, large cell; NOS, not otherwise specified; SC, small-cell lung cancer; SQ, squamous cell.

*Disease extent was defined as follows, according to SEER Extent of Disease – 1988, 3rd edition. Localized: an invasive neoplasm confined entirely to the organ of origin (lung). Regional: a neoplasm that has extended (1) beyond the limits of the organ of origin (lung) directly into surrounding organs or tissues; (2) into regional lymph nodes by way of the lymphatic system; or (3) by a combination of extension and regional lymph nodes. Distant: a neoplasm that has spread to parts of the body remote from the primary tumor either by direct extension or by discontinuous metastasis (eg, implantation or seeding) to distant organs, tissues, or via the lymphatic system to distant lymph nodes.