

# The Prognosis of Pulmonary Sarcomatoid Carcinoma: Development And Validation of A Nomogram Based On SEER

**Yuanyuan Xie**

Wenzhou Medical University First Affiliated Hospital: The First Affiliated Hospital of Wenzhou Medical University

**Zhiyong Lin**

Wenzhou Medical University Second Affiliated Hospital

**Haochun Shi**

Wenzhou Medical University

**Xiang Sun**

Wenzhou Medical University First Affiliated Hospital: The First Affiliated Hospital of Wenzhou Medical University

**Lizhong Gu** (✉ [doctorglz@cumt.edu.cn](mailto:doctorglz@cumt.edu.cn))

Wenzhou Medical University Second Affiliated Hospital <https://orcid.org/0000-0002-9481-1562>

---

## Research Article

**Keywords:** pulmonary sarcomatoid carcinoma, nomogram, overall survival, SEER

**Posted Date:** June 8th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-525695/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background:** The rarity of pulmonary sarcomatoid carcinoma (PSC) and the lack of prospective clinical trials have led to limited knowledge of its clinical characteristics. This study aimed to evaluate the survival and prognostic factors of PSC and to build a nomogram for clinical practice.

**Methods:** Eligible patients diagnosed from 2010 to 2016 were selected from the Surveillance, Epidemiology, and End Results (SEER) database. We compared the clinical characteristics and survival times of PSC patients with those of lung adenocarcinoma (LADC) and lung squamous cell carcinoma (LSCC) patients. We also used univariate and multivariable Cox regression to estimate mortality hazard ratios (HRs) among patients with PSC, while a visual nomogram was established to judge the prognosis. Discrimination, calibration, clinical utility, and reproducibility were validated by Harrell's concordance index (C-index), the area under the curve (AUC), calibration curves, and decision curve analysis (DCA).

**Results:** A total of 400 PSC patients (0.42%) were identified in the SEER database, while 58474 and 33637 patients were diagnosed with LADC and LSCC, respectively. Age, T stage, grade, surgery and radiation were shown to be significant prognostic factors in the Cox regression analyses and were included in the nomogram as predictors. The C-index of the nomogram in the validation set was 0.759. The AUC also demonstrated the good performance of the nomogram, and DCA demonstrated its good clinical applicability.

**Conclusion:** We established a novel nomogram to predict the prognosis of PSC, which can help clinicians make tailored decisions and adjust follow-up management strategies and can provide accurate and individualized survival predictions.

## Introduction

Pulmonary sarcomatoid carcinoma (PSC) is a rare and aggressive type of lung cancer, accounting for approximately 0.1% to 0.4% of all lung malignancies (1) and approximately 1% of non-small cell lung cancers (NSCLCs) (2). According to the World Health Organization (WHO) classification in 2015, PSCs can be categorized into 5 subgroups: pleomorphic carcinoma, spindle cell carcinoma, giant cell carcinoma, carcinosarcoma, and pulmonary blastoma(3). PSCs have poorly differentiated pathological features, reflecting their unfavorable prognosis, and tend to have a poor prognosis irrespective of the stage of disease, with a median overall survival (OS) ranging from 8-19 months and a 5-year survival rate ranging from 17-29% (4-6), presenting relatively worse clinical outcomes than NSCLC (7). Because of its inherent rarity, there have been limited systematic reports describing its clinical manifestation, treatment and prognosis, and there were no significant differences in malignant behavior compared with NSCLCs. The patterns and incidence of distant metastasis of PSCs, especially the brain, bone, adrenal gland, lung and liver, are considered to be found in the majority of diagnosed patients (8, 9). Surgical resection, when feasible, is considered the main treatment, and the role of chemoradiotherapy is controversial and rarely effective (10).

The prognosis of these patients remains controversial. To date, the most important prognostic factors for PSC have been proposed by the American Joint Committee on Cancer (AJCC) staging system, which is based on the depth of tumor invasion (T), the number of metastatic lymph nodes (N), and distant metastases (M). According to the AJCC staging system, age and therapeutic approaches were not involved, which might affect survival. Thus, we need a dedicated clinical model to accurately estimate the prognostic outcome for patients with PSC. Recently, the nomogram, which is specific to an individual patient, has been considered a commonly viable predictive model by graphical calculation, incorporating all the effective clinical features, and has been widely used to estimate the prognosis in multiple malignancies (11). Therefore, we constructed a survival nomogram to predict the prognosis of PSC patients.

The Surveillance, Epidemiology, and End Results (SEER) database is a U.S. population-based cancer database that provides profusion and integral information for different cancers from 20 cancer registries covering ~28% of the population (12). Given the rarity of PSCs and the resulting prognostic data, we sought to better understand PSCs by the SEER database. To meet current clinical needs, our study sought to establish and validate a nomogram to predict the survival of PSC patients based on SEER data.

## Methods

### Patients and study design

Information on patients diagnosed with PSC, LADC and LSCC from 2010 to 2016 was retrieved from the SEER 21 registry database using SEER\*Stat 8.3.5 software. Among 134903 patients diagnosed with PSC (ICD-O-3 8022/3, 8031/3, 8032/3, 8980/3), LADC (ICD-O-3 8040/3, 8144/3, 8230/3, 8250/3, 8253/3, 8254/3, 8333/3, 8480/3, 8551/3) and LSCC (ICD-O-3 8070/3, 8071/3, 8072/3, 8083/3) were identified from the SEER database. Patients with unknown information about race, tumor stage, surgery status, metastatic status and follow-up information were excluded. Finally, a total of 400 cases with PSC, 58474 cases with LADC and 33637 cases with LSCC without missing values were assigned as training sets (Fig. 1).

Demographic or clinical information, including age, sex, race, pathological characteristics of tumors, treatments and distant metastasis (bone, brain, lung and liver), was extracted from the SEER database. The clinical characteristics of the patients included age (<50, 50-69 and  $\geq$ 70 years), sex (male, female), and race (white, black, others). The pathological characteristics of the tumors included the primary site (main bronchus, upper lobe, middle lobe, lower lobe, others) and T stage (T0, T1, T2, T3, T4), N stage (N0, N1, N2, N3), and disease stage according to the 7th edition of the AJCC classification (I, II, III, IV). The treatments included surgery (yes, no), radiotherapy (yes, no/unknown) and chemotherapy (yes, no/unknown).

The primary outcome was overall survival (OS), which was defined as the duration from diagnosis to death, regardless of the cause of death, or was censored at the date of the last follow-up evaluation. The

follow-up time was recorded as the period from the last day of the available survival information or the day of diagnosis to death.

## **Statistical analysis**

Demographic and clinical characteristics of the patients with PSC, LADC and LSCC were compared using Pearson's chi-squared test or Fisher's exact test. OS was estimated using the Kaplan-Meier method by log-rank test with ggplot2, survminer and survival packages in R software version 3.6.0. Adjusted HRs with 95% CIs were calculated using Cox proportional hazards regression models to estimate prognostic factors associated with OS. SPSS (26.0) was used for statistical analysis in this study, and all p values less than 0.05 were statistically significant.

## **Development of the nomogram**

On the basis of the results of the multivariable analyses, a nomogram was constructed that can provide visualized risk prediction using survival. We used univariate and multivariable Cox regression analyses to screen out risk factors and independent prognostic factors for OS, including age, T stage and disease stage according to the 7<sup>th</sup> edition of the AJCC classification, and surgery and radiation, based on the SEER database. Nomograms for 1-, 3- and 5-year OS were constructed on the basis of the identified independent prognostic factors. We then developed a novel visual nomogram.

## **Validation of the nomograms**

The accuracy of the nomogram was assessed by discrimination and calibration evaluation. The consistency index (C-index) and decision curve analyses (DCA) were used to evaluate the predictive accuracy of the nomogram. The C-index was defined as the ratio of all patient pairs predicted to be consistent with the results. Furthermore, 1-, 3-, and 5-year receiver operating characteristic (ROC) curves were used to assess the predictive accuracy of the constructed nomogram. DCA is a method for evaluating alternative diagnostic or prognostic tools and can be used to visualize the clinical consequences of a treatment method (13); thus, DCA was carried out to help us more intuitively understand the value of the nomogram in this study. In addition, to verify the independent discriminatory ability of the nomogram, the patients were divided into high-, moderate-, and low-risk groups according to their total risk score by X-tile. Survival curves for different risk groups were generated using Kaplan–Meier analyses and were detected using the log-rank test. Based on the above, we further plotted clinical impact curves to evaluate the clinical impact of the nomograms to help us more intuitively realized its significant value. The nomogram for predicting OS demonstrated that cost/benefit ratios were lower when the risk threshold was less than 0.3 (Fig. 9B)

# **Results**

## **Patient characteristics**

The demographic or clinical characteristics of the patients are presented in Table 1. We identified 400 eligible patients diagnosed with PSC from 2010 to 2016 and 58474 and 33637 patients diagnosed with LADC and LSCC, respectively, in the SEER database. Among the patients with PSC, 31 (7.75%), 194 (48.50%) and 175 (43.75%) patients were aged <50, 51-69 and  $\geq 70$  years old at diagnosis, respectively. By comparison, more LADC patients were 51-69 years old (51.95%,  $P < 0.01$ ), and more LSCC patients were aged  $\geq 70$  years (51.23%,  $P < 0.01$ ) at diagnosis. Among PSC patients, 58.25% were male, and 41.75% were female, compared to 46.79% of patients who were male and 53.21% of patients who were female among LADC patients and 62.10% who were male and 37.90% who were female among LSCC patients. The majority of PSC patients were white, with tumors located in the upper lobe, and their disease stage was more advanced. In detail, more PSC patients than LSCC patients had stage T3 disease (33.25% vs 23.81%,  $P < 0.01$ ). Moreover, compared to LADC patients, more PSC patients had stage T3 disease, and fewer patients had stage T1 disease (33.25% vs 19.86%, 12.25% vs 28.08%,  $P < 0.01$ ). PSC patients were more frequently diagnosed with N0 and N2 stage disease and less frequently diagnosed with N1 and N2 stage disease. Significantly more PSC patients than LADC patients had stage II and III disease (20.50% vs 8.76%, 22.75% vs 17.74%;  $P < 0.01$ ), and more PSC patients than LSCC patients had stage II and IV disease (20.50% vs 14.66%, 40.25% vs 30.06%;  $P < 0.001$ ). Approximately half of the PSC patients underwent surgery and chemotherapy, while one-third received radiation. Fewer LADC and LSCC patients were treated with surgery than PSC patients, but they received more radiotherapy.

**Table 1.** Demographic and clinical characteristics of patients with PSC, LADC and LSCC.

Characteristics	PSC (n = 400)	LADC (n = 58474)	LSCC (n = 33637)	Total (n =92511)	P-value*	P-value**
Age					<0.01	<0.01
<50	31	3140	814	3985		
50-69	194	30380	15591	46165		
≥70	175	24954	17232	42361		
Sex					<0.01	0.11
Male	233	27361	20888	48482		
Female	167	31113	12749	44029		
Race					0.03	<0.01
White	309	42628	27992	70929		
Black	64	7223	3935	11222		
Others	27	5623	1710	7360		
Primary Site					0.11	0.01
Main bronchus	11	1262	1825	3098		
Upper lobe	253	33489	18819	52561		
Middle lobe	16	2781	1239	4036		
Lower lobe	95	16247	9747	26089		
Others	25	4695	2007	6727		
T stage					<0.01	<0.01
T0	2	328	85	415		
T1	49	16421	6465	22935		
T2	127	17826	11516	29469		
T3	133	11611	8010	19754		
T4	89	12288	7561	19938		
N stage					<0.01	0.03
N0	201	26843	15340	27044		
N1	53	5106	3572	8731		
N2	114	18742	11325	30181		

N3	32	7783	3400	11215		
AJCC 7th Stage					<0.01	<0.01
I	66	15693	8382	24141		
II	82	5123	4930	10135		
III	91	10373	10214	20678		
IV	161	27285	10111	37557		
Surgery					<0.01	<0.01
Yes	197	19846	10149	30192		
No	203	38628	23488	62319		
Radiation					0.12	<0.01
Yes	151	24307	16911	41369		
No/Unknown	249	34167	16726	51142		
Chemotherapy					0.39	0.06
Yes	208	29142	15914	45264		
No/Unknown	192	29332	17723	47247		

Abbreviations: PSC = Pulmonary sarcomatoid carcinoma; LADC = Lung adenocarcinoma; LSCC = Lung squamous cell carcinoma. P-value\*: Comparison between PSC with LADC. P-value\*\*: Comparison between PSC to LSCC.

## Survival

The OS for patients with PSC, LADC and LSCC can be illustrated by Kaplan-Meier plots (Fig. 2). PSC patients exhibited a worse OS than the other patient groups ( $P < 0.001$ , Fig. 3, Table 4). The survival analyses also showed that older patients had much poorer survival than younger patients ( $p = 0.0045$ , Fig. 5A). In addition, the OS of PSC patients with an advanced T stage was significantly shorter than that of PSC patients with a low T stage ( $p < 0.0001$ , Fig. 5B). Moreover, these patients had a better prognosis according to the earlier 7<sup>th</sup> edition of the AJCC guidelines ( $p < 0.0001$ , Fig. 5C). Patients who underwent surgery had better OS than those who did not ( $p < 0.0001$ , Fig. 6A). Similar to surgery, the prognosis of patients receiving radiation was much better than that of patients without radiation ( $p = 0.014$ , Fig. 6B).

**Table 4.** Survival probabilities of each stage in PSC, LADC and LSCC.

Stage	OS	1-Year	3-Year	5-Year	Median	95%CI
All	PSC	42%	27%	21%	9	7.23-10.77
	LADC	60%	37%	27%	20	9.63-20.37
	LSCC	54%	30%	21%	15	14.69-15.31
Stage I	PSC	80%	61%	40%	58	41.67-74.33
	LADC	91%	75%	62%	-	-
	LSCC	83%	58%	44%	50	47.76-52.24
Stage II	PSC	60%	41%	30%	15	1.42-28.58
	LADC	81%	56%	43%	47	44.01-45.00
	LSCC	68%	41%	31%	25	23.58-26.42
Stage III	PSC	48%	28%	26%	12	8.70-15.30
	LADC	67%	36%	23%	23	22.26-23.74
	LSCC	53%	24%	16%	14	13.58-14.42
Stage IV	PSC	15%	6%	6%	3	2.11-3.89
	LADC	36%	12%	6%	8	7.84-8.16
	LSCC	25%	6%	3%	5	4.83-5.17

Abbreviations: PSC = Pulmonary sarcomatoid carcinoma; LADC = Lung adenocarcinoma; LSCC = Lung squamous cell carcinoma.

In the SEER cohort, the median survival times from diagnosis were 9, 20 and 15 months for PSC, LADC and LSCC patients, respectively, and 58 months for PSC patients with stage I disease, 15 months for patients with stage II disease, 12 months for patients with stage III disease and 3 months for patients with stage IV disease. All patients in stage IV had a poor prognosis. Since more than half of stage I LADC patients are still alive after the end of follow-up, the median survival time was not reached.

### Features influencing prognosis

Univariate and multivariate Cox regression models identified the clinicopathological factors that were independently correlated with prognosis in OS for patients with PSC, namely, age at diagnosis ( $P < 0.01$ ), T stage ( $P < 0.01$ ), disease stage according to the 7<sup>th</sup> edition of the AJCC guidelines ( $P < 0.01$ ), surgery ( $P < 0.01$ ) and radiation ( $P < 0.01$ , Table 3). Moreover, the multivariate analysis showed that patients aged 70 years or older had a worse OS than patients younger than 50 years (HR: 2.07, 95% CI: 1.23–3.50). Advanced T stage and advanced disease stage according to the 7<sup>th</sup> edition of the AJCC guidelines were risk factors for PSC patients. Patients who both underwent surgery and received radiation had a better

prognosis than those who did not (HR: 2.29, 95% CI: 1.64–3.20,  $p < 0.01$ ; HR: 1.43, 95% CI: 1.10–1.86,  $p < 0.01$ ).

**Table 3.** Univariate and multivariate Cox regression analysis of prognostic factors in overall PSC patient cohort.

Characteristics	Univariate analysis		Multivariate analysis			
	HR	95%CI	P-value	HR	95%CI	P-value
Age			<0.01			<0.01
<50	Ref.			Ref.		
50-69	1.09	0.68-1.77	0.72	1.35	0.81-2.27	0.25
≥70	1.56	0.97-2.52	0.07	2.07	1.23-3.50	0.01
Sex				NA		
Male	Ref.					
Female	0.85	0.67-1.07	0.17			
Race			0.23	NA		
White	Ref.					
Black	1.29	0.95-1.75	0.10			
Others	0.95	0.59-1.51	0.81			
Primary Site			0.08	NA		
Main bronchus	Ref.					
Upper lobe	0.53	0.27-1.03	0.06			
Middle lobe	0.54	0.23-1.29	0.16			
Lower lobe	0.61	0.31-1.23	0.17			
Others	0.89	0.41-1.94	0.77			
T stage			<0.01			<0.01
T0	Ref.			Ref.		
T1	0.71	0.10-5.32	0.74	1.14	0.15-8.76	0.90
T2	1.67	0.23-12.01	0.61	2.81	0.38-20.77	0.31
T3	2.64	0.37-19.00	0.33	3.98	0.54-29.09	0.17
T4	3.32	0.46-23.85	0.23	3.81	0.52-27.87	0.19
N stage			<0.01			0.87
N0	Ref.			Ref.		
N1	1.35	0.95-1.92	0.10	0.89	0.60-1.30	0.53
N2	1.83	1.41-2.38	<0.01	0.89	0.63-1.26	0.51

N3	1.80	1.17-2.77	<0.01	0.41	0.51-1.38	0.50
AJCC 7th Stage			<0.01			<0.01
I	Ref.			Ref.		
II	1.71	1.08-2.70	0.02	0.98	0.59-1.64	0.95
III	2.25	1.46-3.49	<0.01	1.42	0.79-2.55	0.25
IV	5.67	3.78-8.51	<0.01	2.29	1.64-3.20	<0.01
Surgery						
Yes	Ref.			Ref.		
No	3.28	2.57-4.19	<0.01	2.29	1.64-3.20	<0.01
Radiation						
Yes	Ref.			Ref.		
No/Unknown	0.76	0.60-0.95	0.02	1.43	1.10-1.86	<0.01
Chemotherapy				NA		
Yes	Ref.					
No/Unknown	1.22	0.97-1.53	0.09			

Abbreviations: CI, confidence interval; HR, hazard ratio; PSC = Pulmonary sarcomatoid carcinoma; LADC = Lung adenocarcinoma; LSCC = Lung squamous cell carcinoma.

\*LCNEC = large-cell neuroendocrine carcinoma Ref. = Reference NA = not available

### Difference in metastatic patterns among the three histological types

The common metastatic sites of PSCs were bone (12.8%), lung (12.3%), brain (10.8%) and liver (3.8%) (Table 2, Fig. 4). Among the patients with metastatic disease, PSC patients had significantly fewer bone (P<0.01), liver (P<0.01) and brain (P=0.03) metastases than LADC patients. In addition, LSCC patients had different metastatic patterns than patients with PSC, who had significantly more bone (P=0.02) and brain (P<0.01) metastases.

**Table 2.** Frequencies of each metastasis in PSC,LADC and LSCC.

	PSC n = 400 (%)	LADC n = 58474 (%)	LSCC n = 33637 (%)	P*	P**
Bone metastasis				<0.01	0.02
Yes	51(12.8)	11533(19.7)	3149(9.4)		
No	349(87.3)	46941(80.3)	30488(90.6)		
Brain metastasis				0.03	<0.01
Yes	43(10.8)	8600(14.7)	1661(4.9)		
No	357(89.3)	49874(85.3)	31976(95.1)		
Liver metastasis				<0.01	0.26
Yes	15(3.8)	4278(7.3)	1675(5.0)		
No	385(95.3)	54196(92.7)	31962(95.0)		
Lung metastasis				0.09	0.07
Yes	49(12.3)	8957(15.3)	3230(9.6)		
No	351(87.8)	49517(84.7)	30407(90.4)		

Abbreviations: PSC = Pulmonary sarcomatoid carcinoma; LADC = Lung adenocarcinoma; LSCC = Lung squamous cell carcinoma. P-value\*: Comparison between PSC with LADC. P-value\*\*: Comparison between PSC to LSCC.

### Development of the nomogram and risk stratification model

Based on the selected significant parameters via the Cox regression model, a prognostic nomogram was established for OS (Fig. 7). The nomogram showed that T stage contributed the most to prognosis (T4: score 100; T3: score 96; T2: score 72, T1: score 3), followed by disease stage based on the 7th edition of the AJCC guidelines (IV: score 75, III: score 22, II: score 5), surgery (no surgery: score 68), age ( $\geq 70$ : score 55; 50–69: score 25), and radiation (no/unknown radiation: score 32). After summing the scores associated with each variable and projecting the total scores to the bottom scores, a patient's probability of individual survival can easily be calculated at each time point. Furthermore, a risk stratification system was established based on each patient's total scores from the nomogram to stratify all patients into three risk subgroups: low-risk group (total score <200), intermediate-risk group (total score 200-249), and high-risk group (total score  $\geq 250$ ). In addition, the median survival times of the low-, intermediate- and high-risk groups were 38.0, 6.0 and 4.0 months, respectively, and Kaplan-Meier methods indicated that the risk stratification model could differentiate survival outcomes among the three groups ( $P < 0.0001$ , Fig. 10).

### Validation and clinical performance of the nomogram

For PSC patients, the C-index of the nomogram in the validation set was 0.759 (95% CI 0.738–0.786). To confirm that the nomogram could effectively predict the prognosis of PSC patients, time-dependent ROC analyses at 1, 3 and 5 years were conducted. The 1-, 3-, and 5-year AUC values of the nomogram for the prediction of OS were 0.827, 0.826, and 0.8 (Fig. 8A-C), respectively. Both the C-indexes and AUC values suggested that these models made accurate predictions and had good discriminative abilities. The calibration curves for the probability of 1-, 3- and 5-year OS demonstrated good consistency between the nomogram prediction and actual survival in the training cohort (Fig. 8D-F).

The DCA results of the nomogram demonstrated the good clinical applicability of the nomogram in terms of 1-, 3-, and 5-year patient survival (Fig. 9A). Based on the above, we further plotted the clinical impact curve to evaluate the clinical impact of the nomogram to help us more intuitively realize its significance. The nomogram for predicting OS demonstrated that cost/benefit ratios were lower than those when the risk threshold was less than 0.3 (Fig. 9B).

## Discussion

To date, few systematic reports have focused on PSC due to the inherent rarity of this disease. We reviewed a large national database for the clinical characteristics and outcomes of this tumor. PSCs are mostly larger than 3 cm, with many of them being massive in size, and they most often present as solitary masses in the upper lobes (9). Similar to NSCLC patients, PSC patients of an advanced age had worse survival (9, 14, 15). Our analysis confirmed this; we observed a higher disease prevalence in the upper lobe, and the T stage was found to be the crucial prognostic factor for PSC. In addition, an elderly age has a significant influence on prognosis. To our knowledge, this is the first study to describe the differences between PSC and other NSCLCs in terms of distant metastasis. For metastatic patients, the most common metastatic site was bone, followed by lung, brain, and liver for PSC and LADC patients, while for LSCC patients, the most common site was lung, followed by bone, liver, and brain.

In our study, we showed that PSC is an aggressive, uncommon variant of NSCLC with worse outcomes than LADC and LSCC. The median survival time of PSC patients was 9 months less than that of patients with LADC and LSCC; the 1-year survival rate was 42%, and the 3- and 5-year survival rates were 27% and 21%, respectively. Prior case series have reported a short median survival, ranging from 8 to 19 months (4-7, 9, 16-19) for PSC patients, which is inferior to other NSCLCs. Moreover, PSC patients present with a more advanced stage and worse survival outcomes than LADC and LSCC patients. Patients have typically already reached moderate to advanced disease stages at the time of diagnosis (19). Similarly, most patients in our study were diagnosed with stage III/IV disease according to the 7<sup>th</sup> edition of the AJCC classification, reflecting worse survival.

In this study, both surgery ( $p < 0.01$ ) and radiation ( $p < 0.01$ ) had a significant impact on prognosis. Early surgery is the preferred treatment for PSC. Because of the lack of specific radiologic characteristics and its highly aggressive nature, patients are diagnosed with PSC in advanced disease stages, and surgery is not a viable treatment. For earlier-stage PSCs, surgical resection is an effective treatment and provides

adequate control (20-22), although these patients tend to experience relapse even after complete resection (23). Radiation therapy improved the 5-year survival rate among patients receiving adjuvant radiation (6). However, another study reported a decreased survival rate for patients receiving radiation (24). Of note, the radiation method was shown to be an independent prognostic factor for PSC in our study. Drug resistance can occur with chemotherapy in a short time, resulting in tumor recurrence (25). Because of the poor prognosis of PSC and considering that there is not enough evidence to show the validity of chemotherapy, the identification of an effective chemotherapy regimen is important, and optimal treatments for PSC still need to be determined (26). However, other studies have confirmed that an ordinary NSCLC chemotherapy response was favorable (21). Due to the small number of patients and individual differences, these data are not sufficient evidence that chemotherapy is ineffective. Currently, there are no standardized management protocols for PSC, and surgical resection for localized PSC is the preferred standard treatment by most guidelines, which is consistent with our study. Although our study implies that chemotherapy should not be routinely delivered and radiation could be selected for palliative antitumor therapy, these findings are still controversial and need future evaluation by randomized controlled trials, as with prior case series.

For many solid tumors, the 7<sup>th</sup> edition of the AJCC staging system provides the most prognostic value. However, some important clinical parameters influencing OS, such as age, sex, race, and treatment, are not included in these guidelines. We identified that age, T stage and disease stage according to the 7<sup>th</sup> edition of the AJCC classification, and surgery and radiation were significantly associated with OS in the SEER database. We therefore implemented a more comprehensive prognostic model in the form of a nomogram. This nomogram includes not only the AJCC staging system but also systemic demographics and other important clinical parameters, which are easily obtained and collected through historical clinical records. Moreover, the nomogram achieved satisfactory accuracy and good reliability and reproducibility, as indicated by the C-index, ROC curve, and DCA. The obtained results suggest that we successfully constructed a reliable nomogram for predicting the 1-, 3-, and 5-year OS of PSC patients, since the nomogram validation demonstrated favorable discrimination and calibration. The clinical value of this nomogram could be feasibly applied to predict the survival probability of each individual patient. A risk stratification model that could accurately stratify PSC patients into three risk subgroups was generated on the basis of each patient's total scores from the nomogram, and the survival benefits were analyzed for each risk group. Furthermore, high-risk patients were recognized by the risk stratification system as those who might benefit from more intensive therapy (e.g., radiation). In addition, these patients can be provided with more psychological or palliative care and encouraged to participate in clinical trials for novel drugs, such as immune checkpoint inhibitors. Not least is the fact that a patient may be provided with an objective assessment by a doctor regarding the future course of their disease and given standard management.

Although the nomogram shows good performance, this study still has several limitations. First, the SEER database was retrospective, and patients with incomplete data were excluded. Second, some potential prognostic parameters, such as molecular classification, should be evaluated to improve the nomogram

in future work. Third, the majority of enrolled patients were white or black, so the nomogram needs to be validated externally with other cohorts, especially Chinese patients. Fourth, the results of this study would be more meaningful if the nomogram model was externally validated by another real-world, independent, large-scale, high-quality cohort, which would prove that our findings could be more widely acceptable. Thus, larger and prospective studies with more prognostic variables are needed to improve our predictive model. Despite these limitations, our prognostic nomogram is an instructive and efficient model for accurately predicting individual survival outcomes for PSC patients.

## **Conclusion**

In conclusion, using a large population-based cohort from the SEER database, we described the special clinical and prognostic characteristics of PSC. Older age, advanced stage, surgery and radiation were identified as negative prognostic factors of PSC. Based on the clinical risk factors, we constructed and validated a prognostic nomogram for PSC that has relatively good accuracy. Moreover, the validation process indicated that the current nomogram provides more efficient OS predictions for PSC patients than other currently available methods. We believe that this simple nomogram could be an easy-to-use and beneficial tool for clinicians to promote personalized survival predictions and to guide clinical decision-making for PSC patients.

## **Declarations**

### **Ethical conduct of research**

The authors state that this article does not contain any studies with human participants or animals so exempt from institutional review board approval. Informed consent from study participants was not required as this was a retrospective analysis of an existing database.

### **Consent for publication**

Not applicable.

### **Availability of data and materials**

The datasets used and/or analyzed during the current study are available from the Surveillance, Epidemiology, and End Results (SEER) database (<http://seer.cancer.gov/data/sample-dua.html>).

### **Competing interests**

We declare that we have no conflict of interest.

### **Financial & competing interests disclosure**

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

## Author Contributions

Yuanyuan Xie and Lizhong Gu participated in the overall design of the study, interpretation of results, and drafting of the manuscript. Zhiyong Lin contributed to statistical analysis. Sun Xiang and Haochun Shi contributed to study design and data collection. All authors read and approved the final paper.

## Acknowledgments

The authors thank the staff at the SEER, and all the patients who participated in the study.

## References

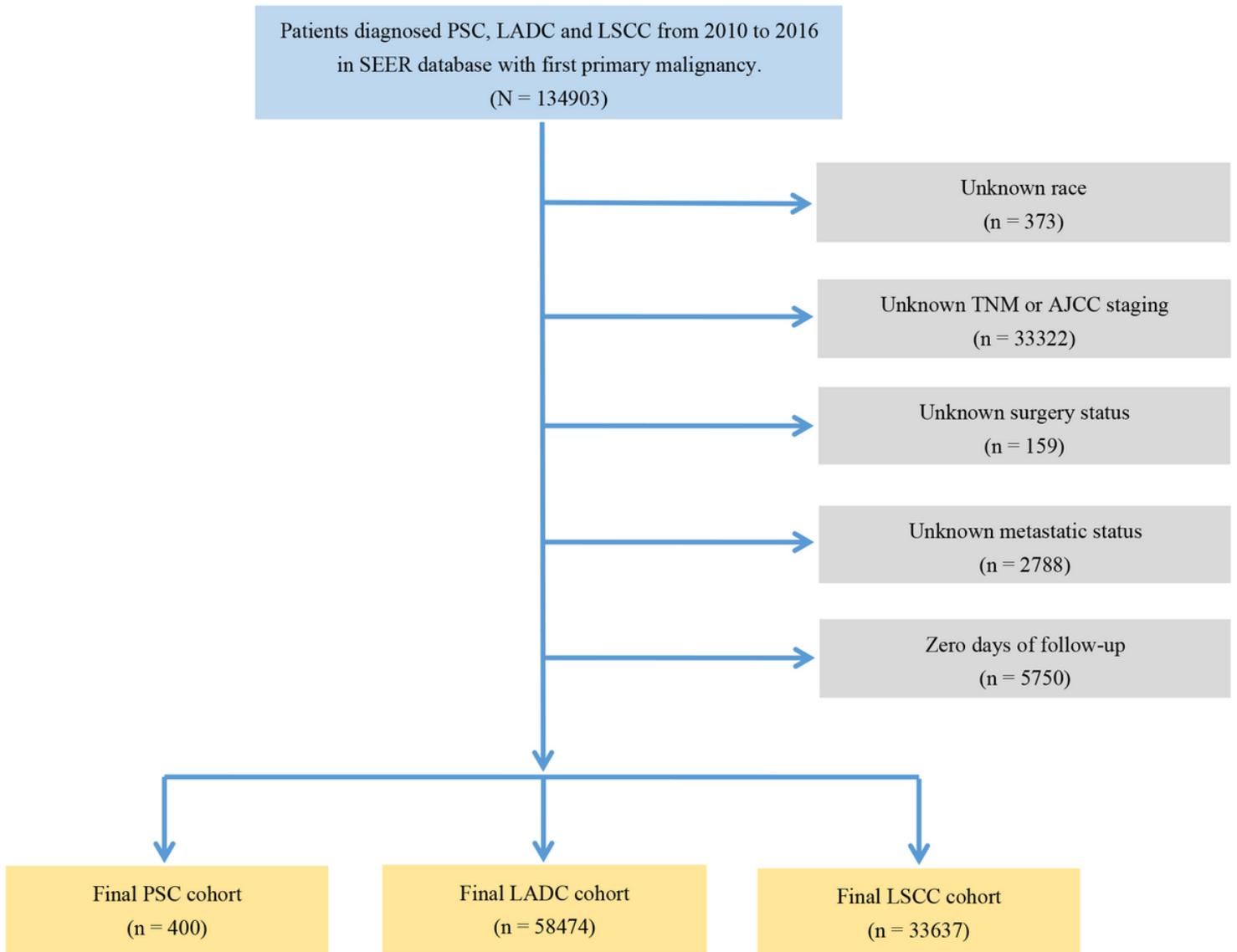
1. Travis WD. Sarcomatoid neoplasms of the lung and pleura. *Archives of pathology & laboratory medicine*.2010;134(11):1645-58.
2. Travis WD, Travis LB, Devesa SS. Lung cancer. *Cancer*. 1995;75:191-202.
3. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, et al. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. *Journal of thoracic oncology: official publication of the International Association for the Study of Lung Cancer*. 2015;10(9):1243-60.
4. Huang SY, Shen SJ, Li XY. Pulmonary sarcomatoid carcinoma: a clinicopathologic study and prognostic analysis of 51 cases. *World journal of surgical oncology*. 2013;11:252.
5. Lin YB, Yang H, Cai QQ, Wang DF, Rao HL. Characteristics and Prognostic Analysis of 69 Patients With Pulmonary Sarcomatoid Carcinoma. 2016;39(3):215-22.
6. Sun L, Dai J, Chen Y, Duan L, Zhang P. Pulmonary Sarcomatoid Carcinoma: Experience From SEER Database and Shanghai Pulmonary Hospital. *The Annals of thoracic surgery*. 2020;110(2).
7. Sai Y, Caty L, Pine M, Adem S, Bogner P, Miller A, et al. Outcomes of sarcomatoid carcinoma of the lung: A Surveillance, Epidemiology, and End Results database analysis. 2012;152(3):397-402.
8. Berho M, Moran CA, Suster S. Malignant mixed epithelial/mesenchymal neoplasms of the lung. *Seminars in diagnostic pathology*.1995;12(2):123.
9. Mochizuki T, Ishii G, Nagai K, Yoshida J, Nishimura M, Mizuno T, et al. Pleomorphic carcinoma of the lung: clinicopathologic characteristics of 70 cases. *The American journal of surgical pathology*. 2008;32(11):1727-35.
10. Chan A, Tong J, Kwan J, Chow C, Chung LY, Chau SL, et al. Assessment of programmed cell death ligand-1 expression by 4 diagnostic assays and its clinicopathological correlation in a large cohort of

- surgical resected non-small cell lung carcinoma. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc.* 2018;31(9):1381-90.
11. Balachandran VP, Gonen M, Smith JJ, Dematteo RP. Nomograms in oncology: more than meets the eye. *The Lancet Oncology.* 2015;16(4):e173-80.
  12. NCI SEER. Number of Persons by Race and Hispanic Ethnicity for SEER Participants (2010 Census Data 1).
  13. Vickers AJ, Elkin EB. Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. *Med Decis Making.* 2006;26(6):565-74.
  14. Aziz H, Rahouma M. Prognostic clinico-pathological features of 99 cases advanced non-small cell lung cancer—Egyptian National Cancer Institute. *Adv Lung Cancer.* 2016;4:29.
  15. Pelosi G, Sonzogni A, Pas TD, Galetta D, Veronesi G, Spaggiari L, et al. Review Article: Pulmonary Sarcomatoid Carcinomas: A Practical Overview. *Int J Surg Pathol.* 2010;18(2):103-20.
  16. Martin LW, Correa AM, Ordonez NG, Roth JA, Swisher SG, Vaporciyan AA, et al. Sarcomatoid Carcinoma of the Lung: A Predictor of Poor Prognosis. *Ann Thorac Surg* 2007;84(3):973-80.
  17. Pelosi G, Gasparini P, Cavazza A, Rossi G, Graziano P, Barbareschi M, et al. Multiparametric molecular characterization of pulmonary sarcomatoid carcinoma reveals a nonrandom amplification of anaplastic lymphoma kinase (ALK) gene %J *Lung Cancer.* 2012;77(3):507-14.
  18. Rossi G, Cavazza A, Sturm N, Migaldi M, Facciolo N, Longo L, et al. Pulmonary carcinomas with pleomorphic, sarcomatoid, or sarcomatous elements: a clinicopathologic and immunohistochemical study of 75 cases. *Am J Surg Pathol.* 2003;27(3):311-24.
  19. Chen J, He Q, Liu J, Xiao Y, Xiao C, Chen K, et al. CD8+ tumor-infiltrating lymphocytes as a novel prognostic biomarker in lung sarcomatoid carcinoma, a rare subtype of lung cancer. *Cancer Manag Res.* 2018;10: 3505–3511.
  20. Zehani A, Ayadi-Kaddour A, Marghli A, Maamouri H, Mezni FE. Sarcomatoid carcinoma of the lung: Retrospective study of 28 cases. *Annales de pathologie* 2014;34(2):124-9.
  21. Sim JK, Sang MC, Choi JH, Oh JY, Lee SY. Clinical and molecular characteristics of pulmonary sarcomatoid carcinoma. *J Korean Journal of Internal Medicine.* 2018;33(4):737-44.
  22. Sun L, Dai J, Wang X, Jiang G, Diego GR, Song J, et al. Pulmonary carcinosarcoma: analysis from the Surveillance, Epidemiology and End Results database. *Interactive cardiovascular and thoracic surgery* 2019(1):1.
  23. Maneenil, Kunlatida, Xue, Zhiqiang, Liu, Ming, et al. Sarcomatoid Carcinoma of the Lung: The Mayo Clinic Experience in 127 Patients. *Clinical lung cancer* 2018;19(3):e323-e333.
  24. Mohamed R, Mohamed K, Navneet N, Abu N, Sebron H, Benjamin L, et al. Pulmonary sarcomatoid carcinoma: an analysis of a rare cancer from the Surveillance, Epidemiology, and End Results database. 2017;53(4):828-834.
  25. Ge J, Yao B, Huang J, Wu X, Bao H, Ou Q, et al. Molecular genetic characterization reveals linear tumor evolution in a pulmonary sarcomatoid carcinomas patient with a novel PHF20-NTRK1 fusion:

a case report. BMC Cancer 2019;19(1):592.

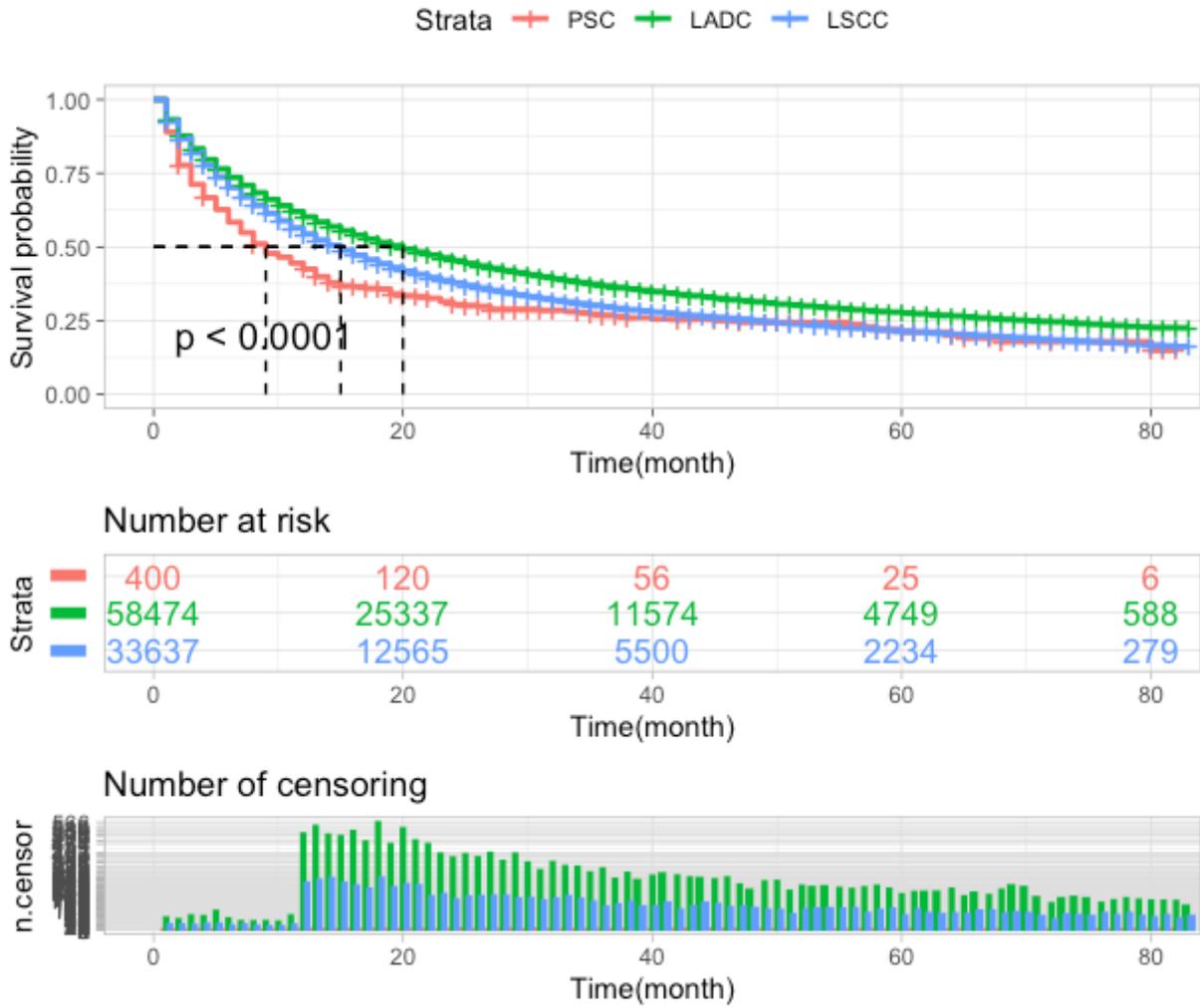
26. Lococo F, Rapicetta C, Cardillo G, Stefani A, Margaritora S, Leuzzi G, et al. Pathologic Findings and Long-Term Results After Surgical Treatment for Pulmonary Sarcomatoid Tumors: A Multicenter Analysis. 2017.

## Figures



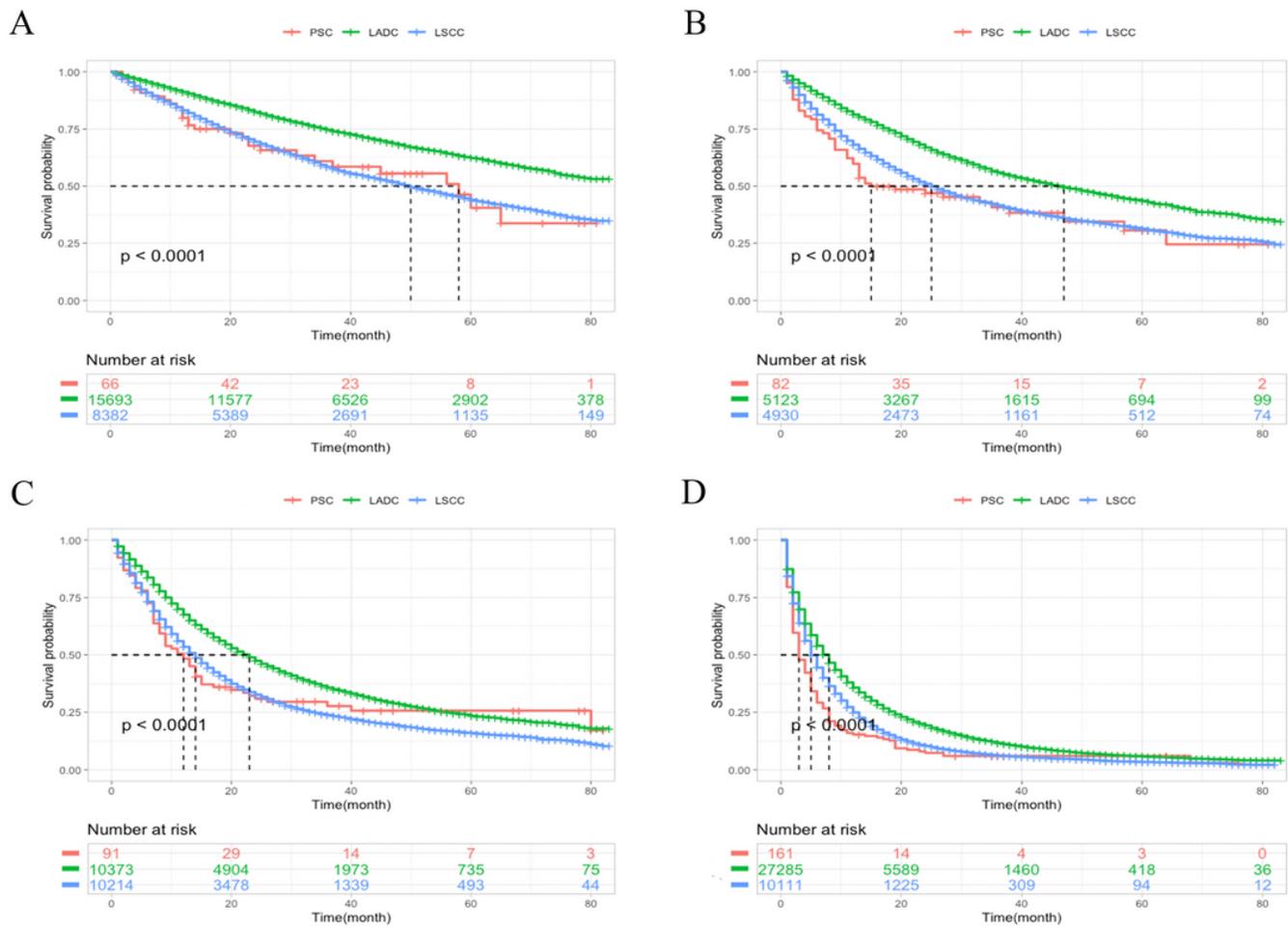
**Figure 1**

Flow chart of patients' cohort selection. PSC: pulmonary sarcomatoid carcinoma; LADC: lung adenocarcinoma; LSCC: lung squamous cell carcinoma.



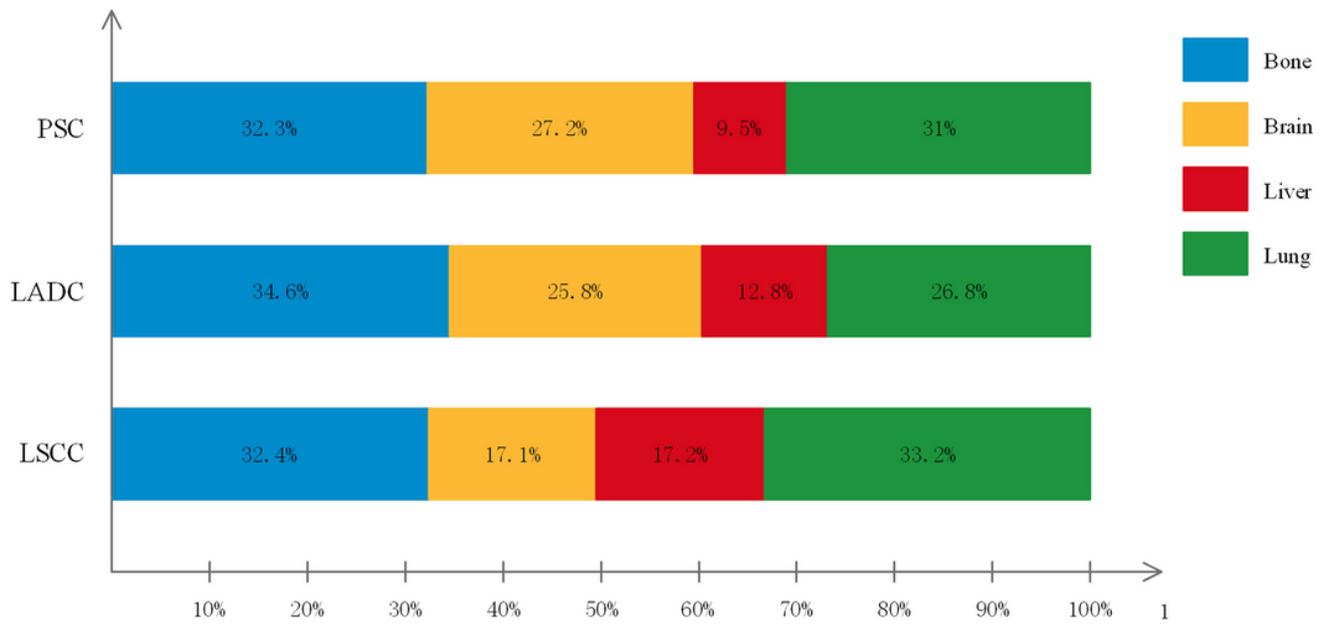
**Figure 2**

The Kaplan–Meier overall survival in PSC, LADC and LSCC. PSC: pulmonary sarcomatoid carcinoma; LADC: lung adenocarcinoma; LSCC: lung squamous cell carcinoma.



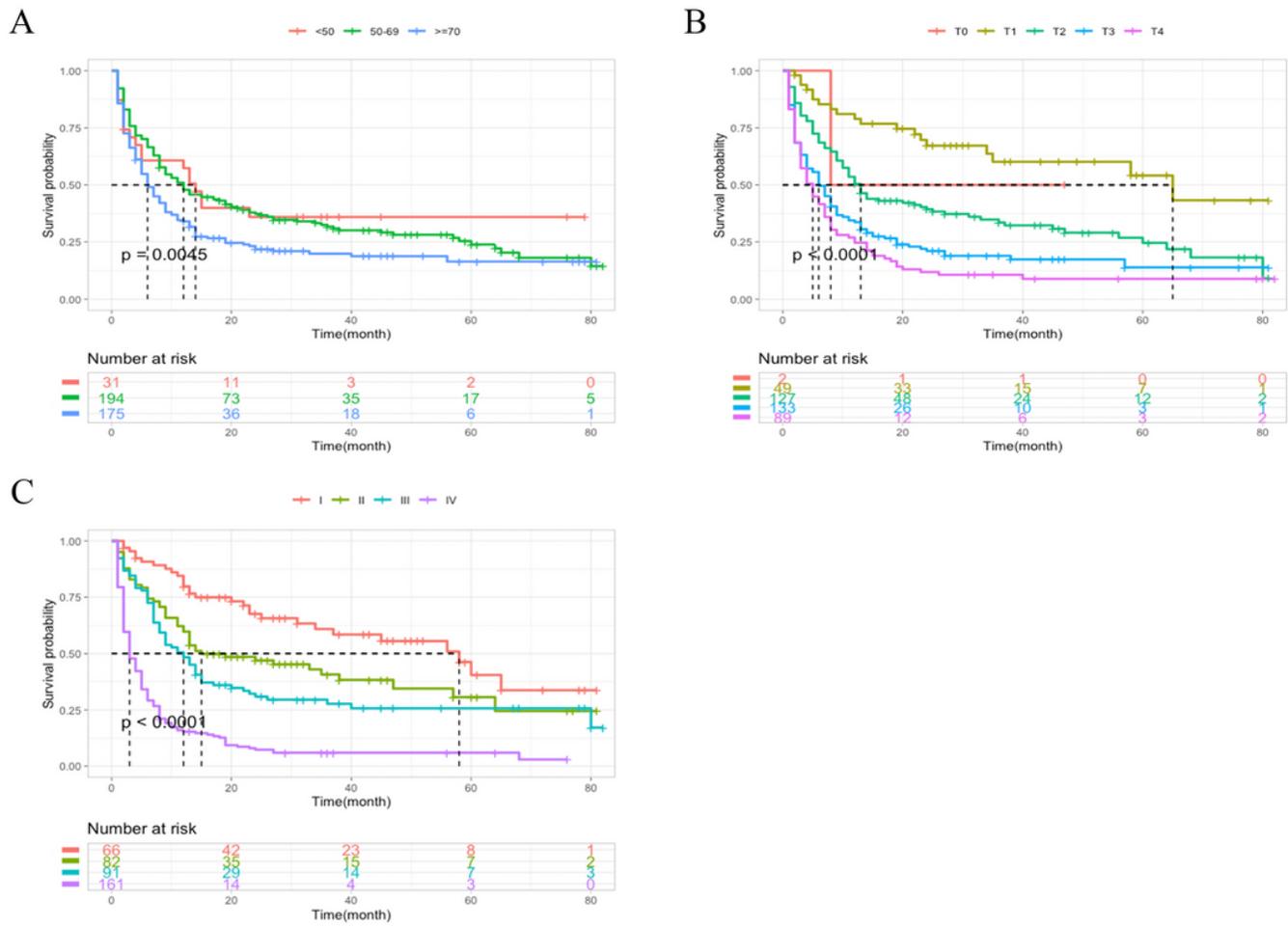
**Figure 3**

The Kaplan–Meier overall survival for different stages according to the 7th edition of the AJCC in PSC, LADC and LSCC. Stage I (A), stage II(B), stage III(C), stage IV(D). PSC: pulmonary sarcomatoid carcinoma; LADC: lung adenocarcinoma; LSCC: lung squamous cell carcinoma.



**Figure 4**

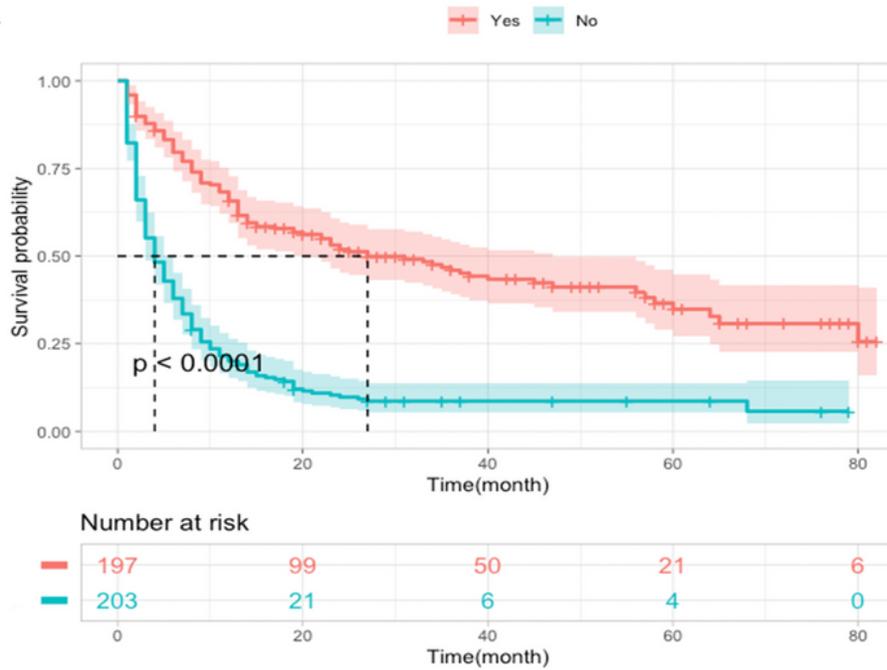
Frequencies of each metastasis in PSC, LADC and LSCC. PSC: pulmonary sarcomatoid carcinoma; LADC: lung adenocarcinoma; LSCC: lung squamous cell carcinoma.



**Figure 5**

Kaplan–Meier analysis of OS in PSC patients stratified by age(A), 7th AJCC T stage(B), disease stage according to the 7th edition of the AJCC(C). OS: overall survival.

A



B

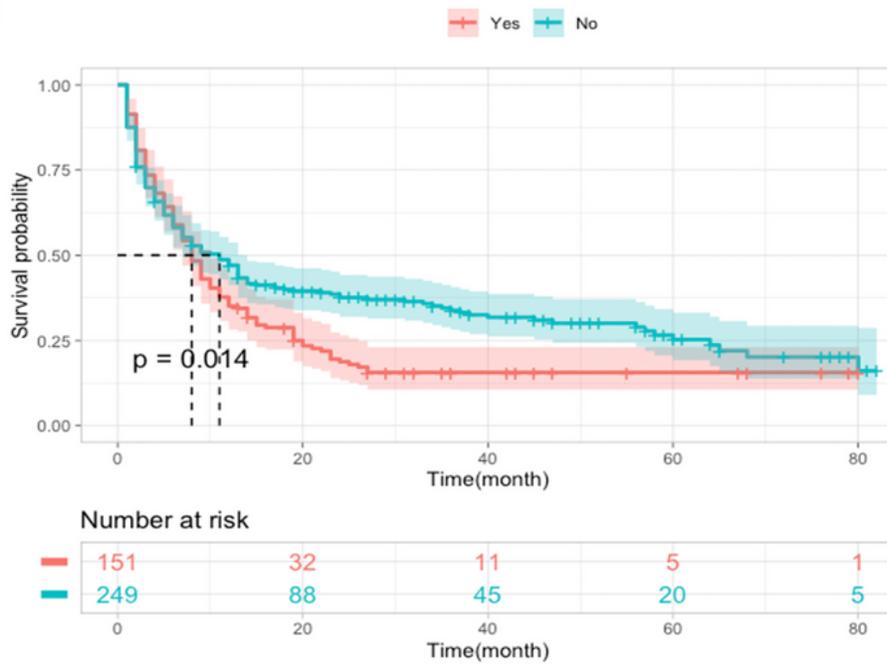
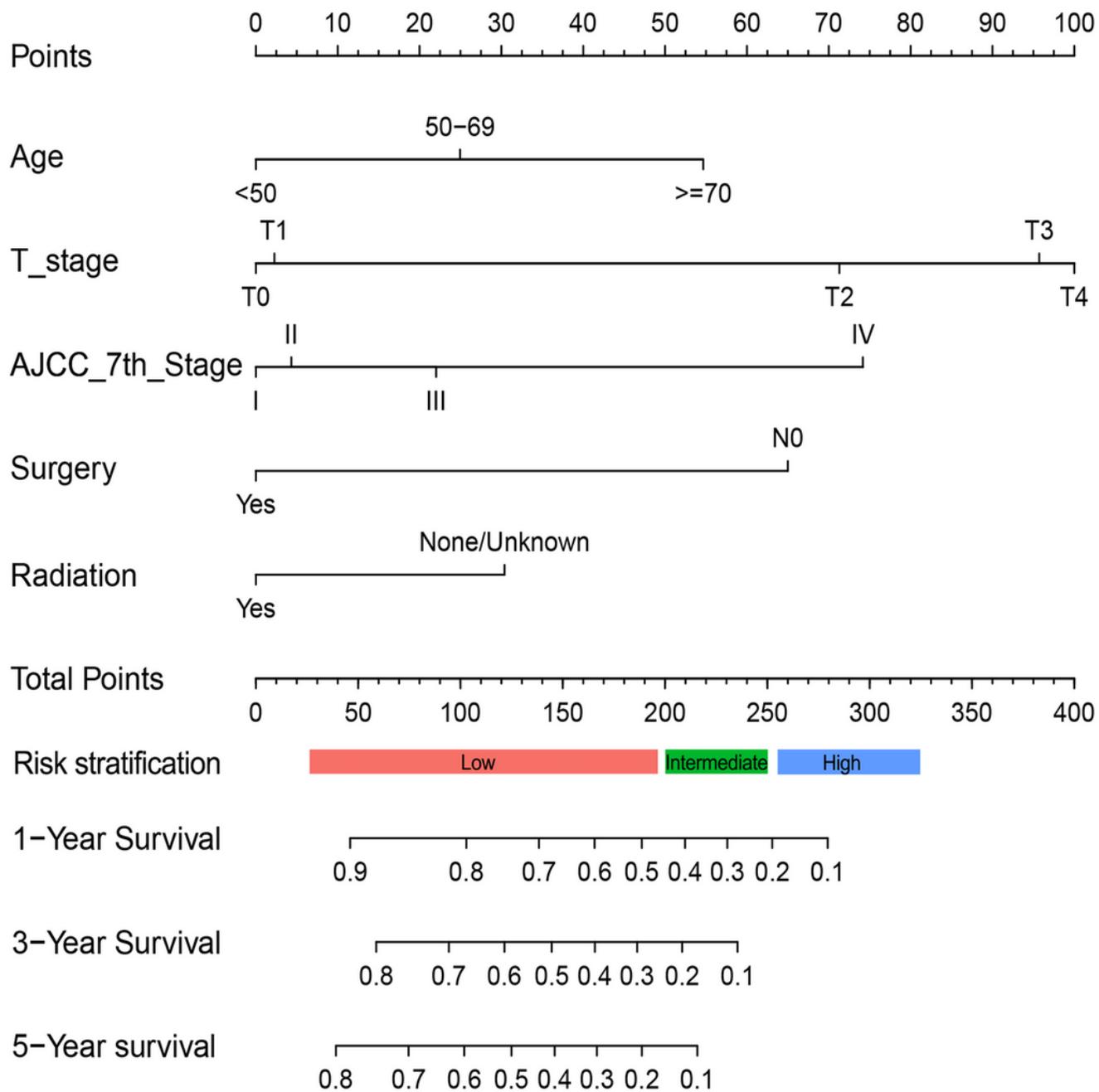


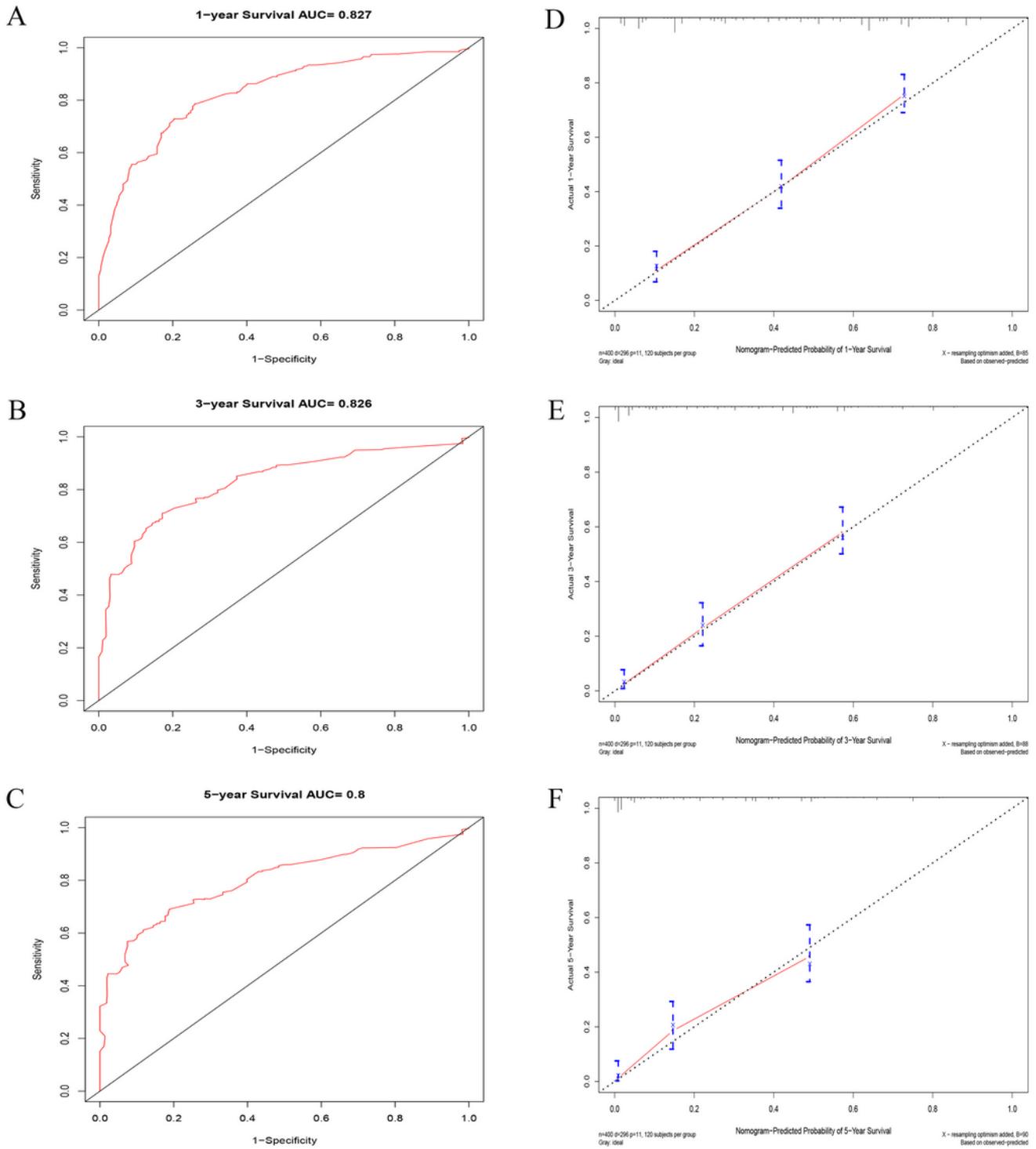
Figure 6

Kaplan–Meier analysis of OS in PSC patients stratified by surgery (D), radiotherapy(E). OS: overall survival.



**Figure 7**

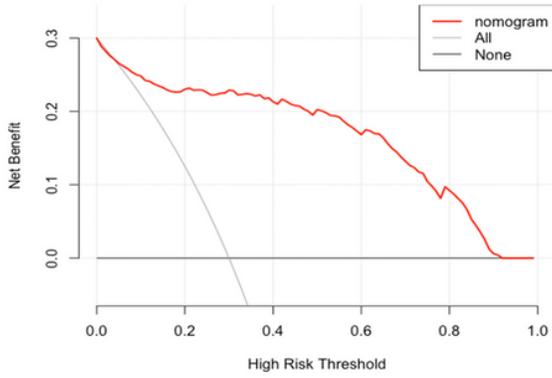
Prognostic nomogram including significant clinical characteristics for 1- year, 3-year, and 5-year OS in PSC patients. OS: overall survival.



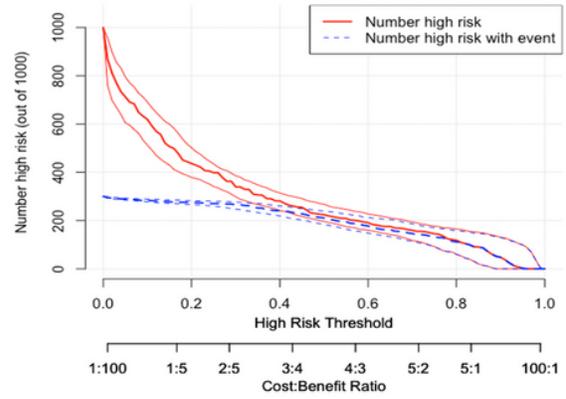
**Figure 8**

ROC curves. The ability of the model to be measured by the C index. 1-year OS (A), 3-year OS(B), 5-year OS(C). Calibration curves of the nomograms predicting overall survival (OS) in PSC patients. 1-year OS(E),3-year OS(D), 5-year OS(F). AUC area under the curve. OS: overall survival.

A

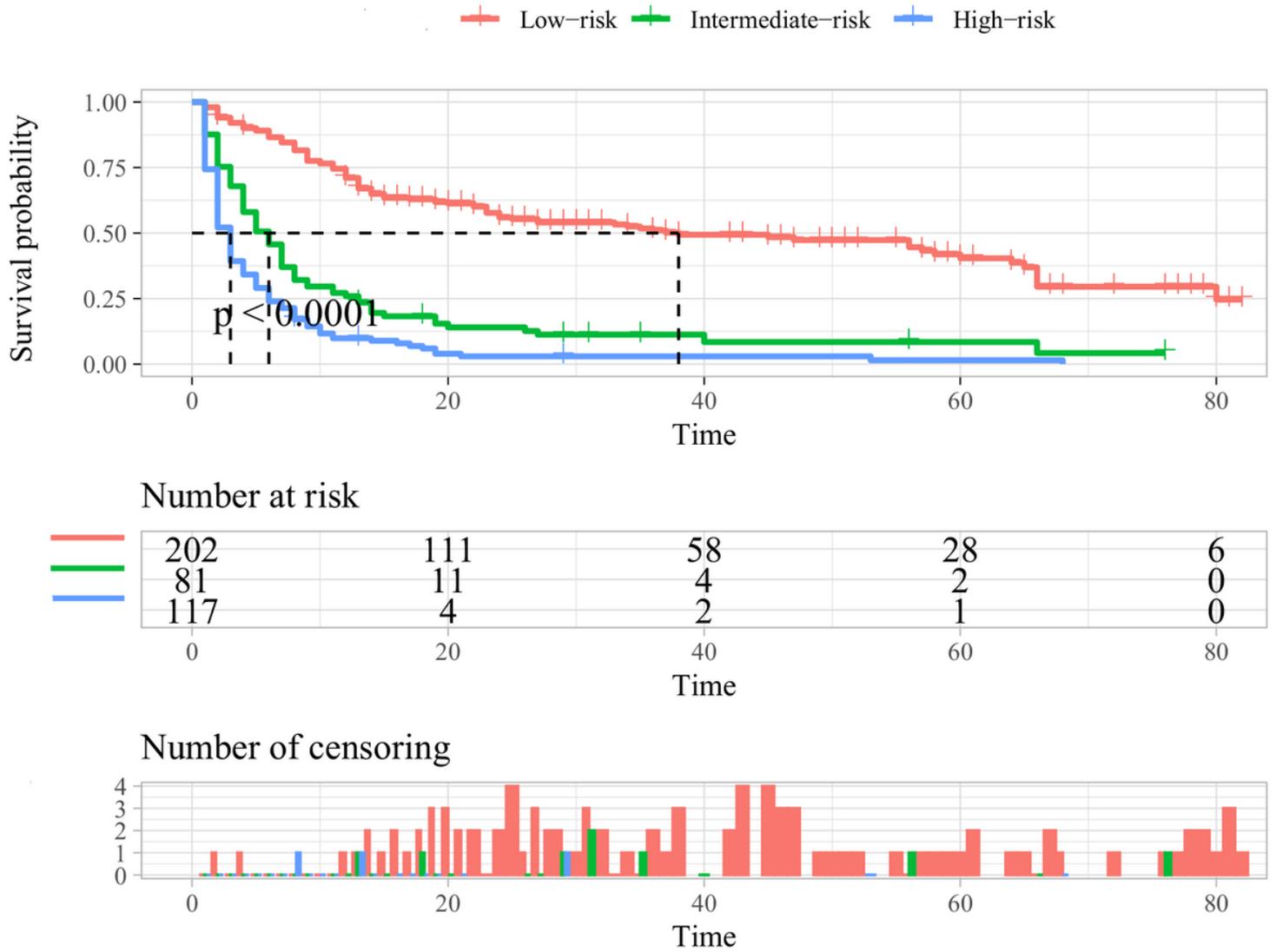


B



**Figure 9**

DCA for nomograms (A). In the figure, x-axis is the threshold probability, y-axis is the net benefit rate. “All” refers to the assumption that all patients reached the endpoint and “none” to the hypothesis that no patients reached the endpoint. Clinical impact curves of the nomograms for OS (B) in PSC patients. The number of high-risk patients and the number of high-risk patients with the outcome are plotted at different threshold probabilities within a given population. OS: overall survival.



**Figure 10**

Survival analysis using the Kaplan-Meier approach identified that a nomogram score.