

Predicting Prognosis In Primary Pulmonary Sarcoma Using A Clinical Nomogram Based On SEER Database

Piao Shen

Guangzhou Medical University Affiliated Cancer Hospital: Guangzhou

Yuzhen Zheng

The Sixth Affiliated Hospital of Sun Yat-sen University

Siyu Zhu

Guangzhou Medical University Affiliated Cancer Hospital: Guangzhou

Xingping Yang

The Sixth Affiliated Hospital of Sun Yat-sen University

Jian Tan

The Sixth Affiliated Hospital of Sun Yat-sen University

Weijie Cai

The Sixth Affiliated Hospital of Sun Yat-sen University

Hong-Ying Liao (✉ liaohy2@mail.sysu.edu.cn)

The Sixth Affiliated Hospital of Sun Yat-sen University

Research Article

Keywords: Nomogram, Overall survival, Primary pulmonary sarcoma, SEER

Posted Date: January 13th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1238447/v1>

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

[Read Full License](#)

Abstract

Background: Primary pulmonary sarcoma (PPS) accounts for less than 1.1% of all pulmonary tumors. Few data outcomes are reported. This study aims to clarify the predictive value of clinicopathologic features on the overall survival (OS) of PPS patients.

Methods: Patients with primary pulmonary sarcoma (PPS) were collected from the Surveillance, Epidemiology, and End Results (SEER) database (from 2000 to 2015) and divided randomly into training and validation cohorts at a ratio of 1:1. Univariate Cox analysis and the least absolute shrinkage and selection operator (LASSO) were implemented to identify prognostic factors related to overall survival of primary pulmonary sarcoma patients. Then, we performed multivariate Cox regression to establish a prognostic factors signature. The Kaplan- Meier (K-M) survival curves and time-dependent receiver operating characteristic (ROC) curves were plotted to estimate the prognostic power of the signature. In addition, multivariate Cox regression screened out independent prognostic factors and constructed a nomogram.

Results: PPS patients with training group were divided into low- and high-risk group based on risk score, and high-risk group had a shorter survival time. The validation group got the same result. ($P < 0.001$). On multivariate analysis of the training cohort, independent factors for survival were marriage, age, sex, grade, operation, metastasis and tumor size, which were all selected into the nomogram. The calibration curve and ROC plots for probability of 3-year and 5-year survival were in accord with prediction by nomogram and actual observation. And the C-index of the nomogram for predicting survival was 0.77 (95% CI, 0.74 to 0.80, $P < 0.05$), which was statistically significant.

Conclusion: We constructed a risk prognosis model based on PPS patients from SEER database. In addition, the construction of nomogram provides one more idea for clinical treatment.

Introduction

Primary lung sarcomas (PPS) are a group of rare non-epithelial malignancies that originate in the lung mesenchymal tissue. It develops from mesenchymal elements of bronchial wall, vessels or pulmonary stroma¹⁻³.

Primary lung sarcoma (PPS) is a rare and aggressive tumor that accounts for only 0.4% – 1.1% of all lung malignancies. Apart from individual case reports, few cohort studies have been reported on primary sarcoma of the lung. Most reports indicated a survival benefit from surgical resection. However, a guideline for eligible treatment of PPS is not yet available. The low incidence and detection rate of this group of tumors brought out the fact that primary sarcomas of the lung still remain a diagnostic and cure challenge⁴⁻⁹.

Therefore, the aim of this study is to examine a large cohort of PPS patients to better characterize the clinicopathologic prognosis factors of PPS that impact survival. Furthermore, prognosis model was

established to predict the probability of survival and provide more comprehensive evidence for early diagnosis and clinical treatment.

Methods

1. Patient Selection and Data Retrieval

Patient information was collected from the SEER database of the National Cancer Institute (<http://seer.cancer.gov/>). After consulting the CS Schema v0204+, we extracted data from 1094 patients with Primary pulmonary sarcoma from 2000 to 2015 using SEER*Stat software (version 8.3.8). The inclusion criteria were as follows: (a) Patients with primary lung sarcoma were confirmed by pathology in SEER database. (b) All patients had tumor-specific death. (c) Follow-up information was complete on marriage, sex, age, race, site, grade, Laterality, stage, operation method, operation, radiotherapy, chemotherapy, neoplasm invasiveness, Lymph node status, metastasis and tumor size. Patients lacking any of this information were excluded from the study cohort.

2. Construction and Validation of Prognostic Model

247 patients with Primary pulmonary sarcoma (PPS) were randomly divided into a training group (N=124) and a test group (N=123) in a ratio of 1 to 1. All preliminary analyses were performed in the training group to construct a feature based on prognostic factors and to validate the feature in the test group. Univariate Cox proportional risk analysis was used to screen for prognostic factors associated with overall survival in patients with primary pulmonary sarcoma, $P < 0.05$ as statistical significance. LASSO analysis is a high-dimensional indicator regression method, eliminating the highly correlated variables and preventing the model from over-fitting. LASSO analysis was used to screen out the critical factors from the general prognostic factors in univariate Cox regression analysis using R with glmnet package (Version 3.0-2, <https://CRAN.R-project.org/package=glmnet>). Then, multivariate Cox regression, stepwise regression, was performed to reduce dimensionality and establish a risk score formula weighted by the corresponding coefficients. The univariate and multivariate Cox regression analysis was performed using survival package (version 2.41-1, <http://bioconductor.org/packages/survival/>) in R language. A risk score was calculated for each patient in the training group based on the above formula. The training and test groups were classified as high-risk and low-risk groups, respectively, based on the median risk value of the training group. Survival difference of five years between the low-risk and high-risk group was assessed by the Kaplan-Meier (K-M) survival analysis using R with survival package (Version 2.41-1, <http://bioconductor.org/packages/survival/>). To evaluate the predictive performance of the five-year survival difference, a time-dependent receiver operating characteristic (ROC) curve was plotted using the R and survivalROC packages (Version 1.0.3, <https://CRAN.R-project.org/package=survivalROC>). Thus, we perform K-M survival and ROC analysis to evaluate predictive accuracy of the prognosis factors in the testing group according to the same cutoff value. The area under the curve (AUC) is served as the evaluation criterion.

3. Construction of nomogram

Based on the results of multivariate Cox regression analysis, we constructed a nomogram of individualized overall survival prediction to estimate 1-, 3-, and 5-year overall survival using R with rms package (Version 5.1-4, <https://CRAN.R-project.org/package=rms>). The performance of the nomogram was measured by concordance index (C-index) and assessed by ROC and calibration curve. The larger the C-index, the more accurate was the survival prediction. During the validation of the nomogram, the total points of each patient in the cohort were calculated according to the established nomogram, then Cox regression in the cohort was performed using the total points as a factor, and finally, the C-index and calibration curve were derived based on the regression analysis. $P < 0.05$ was considered statistically significant.

Results

1. characteristics of the included patients

A total of 247 patients diagnosed with primary sarcoma of the lung met the criteria and were included in the analysis. Among which the male (142 [57.5%]) to female (105[42.5%]) ratio was 1.35:1, and the predominant race was white (215 [87.0%]). To better guide clinical practice, the age was divided into group 1 (age < 20y) , group 2 (20~39y), group 3(40~49y), group 4(50~59y), group 5(60~69y),group 6(70~79y) and group 7($\geq 80y$). While tumor size was classified into group 1 (diameter $\leq 3.0\text{cm}$) , group 2 (3cm \leq diameter $\leq 5\text{ cm}$), group 3(5cm \leq diameter $\leq 7\text{ cm}$),group 4 (7cm \leq diameter $\leq 10\text{ cm}$) and group5 (10cm \leq diameter) . The baseline clinicopathological characteristics, demographic characteristics, and comparable p values are shown in **Table 1**.

2. Identification and selection of independent prognostic factors

First, we randomly divided 247 patients with PPS into two groups: a training group and a test group. Cox regression analysis was performed to identify significant prognostic factors correlated with overall survival (OS). As shown in **Table 2**, univariate regression analysis of the train group revealed that clinicopathological characteristics, including age, grade, stage, lymph node status, and metastasis were prognostic risk factors ($p < 0.01$), whereas, marriage, gender and operation were prognostic protective factors ($p < 0.001$). In order to improve the accuracy of the prediction model, we use Lasso regression to eliminate variables with high correlation, to further reduce the variables. As a result, 7 key prognostic factors were identified from 11 significant clinicopathological characteristics in the univariate regression analysis($P < 0.05$) using lasso regression **figure 2A-B**. Next, multivariate COX regression analysis, stepwise regression and screening, were performed to determine the best risk score formula to predict the overall survival rate of the risk prognosis model (**table 2**). The sites corresponding with these clinicopathological characteristics were marriage, gender, age, grade, operation, metastasis and tumor size.

According to the corresponding coefficients of the prognostic factors β -values, a risk score formula was obtained for predicting prognosis. Risk score = $0.230 \times \beta\text{-value of age} + 0.471 \times \beta\text{-value of grade} + 0.695 \times \beta\text{-value of metastasis} + 0.155 \times \beta\text{-value of tumor size} - 0.710 \times \beta\text{-value of marriage} - 0.553 \times \beta\text{-value of}$

gender - $0.772 \times \beta$ -value of operation. Among these characteristics, age, grade, metastasis and tumor size were negative correlated to overall survival in PPS patients, on the contrary, marriage, gender and operation were positive ones (**figure2C**).

3. Survival analysis

To assess the predictive performance of the selected prognostic factors, patients were divided into high-risk (N=62) and low-risk (N=62) groups using the median risk score of the training group as a threshold. Distribution of risk score, survival status of each patient, and heat map expression of seven prognosis factors were constructed in training group. As the same, they were validated in the test group. K-M survival curves showed patients in high-risk group with shorter OS and low-risk group with longer OS ($P < 0.001$). What's more, the AUC is 0.847 (**figure 3A**). Most importantly, the similar situation was proved in test group and the AUC is 0.813 (**figure 3B**). These results demonstrated that our prognosis factors performed significant sensitivity and accuracy in assessing PPS patients' overall survival.

4. Construction and validation of nomogram

According to the results from multivariate analysis, marriage, gender, age, grade, operation, metastasis and tumor size were significantly related to overall survival of patients with PPS. Through analysis of the above independent and stable prognostic factor, we constructed a visualized nomogram predicting the 1-, 3-, and 5-year survival rates (**figure3C**). The prediction nomogram successfully quantified each indicator's contribution to survival and the concordance index and calibration plots showed its superior predictive capacity. The C-index of the nomogram for predicting OS was 0.77 (95% CI, 0.74 to 0.80, $P < 0.05$), at the same time, 3-year and 5-year Survival AUC were both 0.873 (**figure4A** and **B**). The last but not least, the calibration curve showed good agreement between prediction and observation in the probability of 5-year survival and 3-year survival (**figure 4C** and **D**). All the above results showed that our prognosis prediction model had better stability and reliability predicting overall survival in PPS patients.

Discussion

As we all know, primary lung sarcoma (PPS) is a rare malignancy accounting for less 1% of all lung neoplasm¹⁻³. Since the low incidence and detection of PPS, we are more interested in which clinicopathologic factors are significantly correlated with these patients and desired to develop a predictive nomogram for this cohort. So we conducted a series of analysis of PPS patients from SEER database. Here, we collected 247 PPS patients with complete clinical information performing survival analysis. After assessing univariate and multivariate cox regression, we developed a nomogram containing 7 significant independent prognosis factors, including marriage, gender, age, grade, operation, metastasis and tumor size. All of the above prognostic factors are easy to acquire. In addition, the focus of our conclusions is that patient gender, marriage, and metastasis are the key prognostic factors to be identified in primary pulmonary sarcoma for the first time, providing additional targets for clinical prognostic analysis.

Previous studies have inferred that various significant prognostic factors for patients with PPS, such as age, tumor size, tumor grade, histologic form and treatment type^{10,11}. Similarly, our study also proposed that age, grade and tumor size were significant risk factors, while operation was important protective factors. Matthew Koshy et al. revealed PPSs are associated with a higher rate of nodal metastasis relative to soft-tissue extremity sarcomas¹. Thus, it is recommended that pulmonary sarcoma patients undergo a thorough mediastinal nodal evaluation to rule out locoregional metastasis. In our study, the lymph node status was an independent prognostic factor in univariate Cox regression analysis. The reason why we excluded this factor in the following analysis was avoiding the highly correlated variables between those clinicopathologic factors to improve accuracy of predicting OS. Although the reports on PPS are not abundant, most studies have showed complete surgical resection is important for cure which our results further confirm¹¹⁻¹⁴. Therefore, the benefit of surgery for PPS patients maybe a good choice to a certain degree. Another interesting point is that we identified three different prognostic factors (gender, marriage, and metastasis) comparing with previous study, which means that the objective of clinical prognostic analysis is broadened, allowing physicians to consider outcomes from multiple dimensions. However, it has not been mentioned in previous studies that the prognosis of PPS patients is worse in males and the unmarried than in females and the married. More data studies may be needed to explore the reasons behind this.

Besides the new factors for prognostic analysis, our survival prognostic model also showed good predictive performance. And C-index, AUC and calibration curves all showed that our prognosis model has a high accuracy and stability in 3-year or 5-year OS. C-index of the nomogram for predicting OS was 0.77, Survival AUC were both 0.873. This is one of the most powerful aspects of our model. Overall, 247 samples from SEER database were conducted the first nomogram of PPS survival. We not only confirmed the common prognostic factors mentioned in previous studies, but also verified several different factors, laying a more solid foundation for the prognostic analysis of PPS patients. Moreover, good predictive performance makes it the fact the reliability of this group of randomly grouped data. If the results of our study are confirmed by further clinical practice, it may have a positive impact on the formulation of policies for the diagnosis and prognosis assessment of PPS patients.

Although our model had good performance for predicting survival, there must be some defects. Firstly, it was a retrospective study and the effective samples belong to a medium scale cohort from SEER database. Secondly, some significant prognostic factors, such as some physical indices, resection margin and the level of CEA were not available in the SEER database, resulting in limitation of our study. In addition, many patients in this study were classified as a primary sarcoma that is not otherwise specified which reflects the difficulty in the tissue diagnosis of PPS, and it has been mentioned by Etienne-Mastroianni et al¹². Rapidly evolving diagnostic techniques may allow future studies to further identify the relationship between PPS histology and behavior. Finally, further large prospective studies are needed to confirm the validity of our prognostic model, too.

In conclusion, we firstly conducted and validated a nomogram that has a higher potential to accurately predict OS of PPS patients and could provide some insights into the clinical therapy practice. This result

would assist clinicians in performing personalized clinical treatment and more accurate prognosis estimation.

Declarations

Ethics approval and consent to participate:

Not applicable.

Consent for publication:

Not applicable.

Availability of data and materials

Not applicable.

Competing interests:

The authors declare that they have no competing interests.

Funding:

This study was supported by the Medical Scientific Research Foundation of Guangdong Province of China (Grant No. A2020150).

Authors contributions:

SP prepared and wrote the manuscript, searched literature; ZY planned and designed the study; ZS, TJ and YX collected and interpreted the data; SP, ZY, and CW analyzed the data and organized the logic of the paper. All authors read and approved the final manuscript.

Acknowledgements:

None.

References

1. Spraker MB, Bair E, Bair R, Connell PP, Mahmood U, Koshy M. An analysis of patient characteristics and clinical outcomes in primary pulmonary sarcoma. *Journal of thoracic oncology*. 2013;8(2):147-51.
2. Collaud S, Stork T, Schildhaus HU, Valdivia D, Zaatar M, Dirksen U, et al. Multimodality treatment including surgery for primary pulmonary sarcoma: Size does matter. *Journal of surgical oncology*. 2020;122(3):506-14.

3. Jakubcová T, Jakubec P. [Primary pulmonary sarcomas] *Klinická onkologie : časopis České a Slovenské onkologické společnosti*. 2009;22(4):139-53.
4. Yamada Y, Kaplan T, Soltermann A, Schmitt-Opitz I, Schneiter D, Weder W, et al. Surgical Outcomes and Risk Analysis of Primary Pulmonary Sarcoma. *The Thoracic and cardiovascular surgeon*. 2021;69(1):101-8.
5. Gołota J, Osowiecka K, Orłowski T. Primary pulmonary sarcoma - long-term treatment outcomes and prognostic factors. *Kardiochir Torakochirurgia Pol*. 2018;15(3):162-9.
6. Gołota J, Osowiecka K, Orłowski T. Primary pulmonary sarcoma - treatment outcomes depending on the different types of radical operation. *Kardiochir Torakochirurgia Pol*. 2019;16(1):1-6.
7. Andrushchuk U, Mofeejuddy I, Spirydonau S, Yudina O, Shchatsinka M. Surgical treatment of primary pulmonary sarcoma disseminated to both pulmonary arteries using composite pulmonary trunk and bifurcation of abdominal aorta allograft. *Kardiochir Torakochirurgia Pol*. 2018;15(4):254-7.
8. Liu K, Li W. Analysis of 19 cases of primary pulmonary sarcoma. *Zhongguo Fei Ai Za Zhi*. 2012;15(6):375-80.
9. Arikan I, Barut A, Harma M, Harma MI, Erdem Z, Gezer S. Primary pulmonary sarcoma metastasising to the skin during pregnancy. *J Pak Med Assoc*. 2012;62(12):1342-3.
10. Régnard JF, Icard P, Guibert L, Montpreville VTd, Magdeleinat P, Levasseur P. Prognostic factors and results after surgical treatment of primary sarcomas of the lung. *The Annals of thoracic surgery*. 1999;68(1):227-31.
11. SB K, E B, EJ M, GP N, AE R. Primary pulmonary sarcoma: a clinicopathologic study of 26 cases. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*. 1999;12(12):1124-31.
12. Etienne-Mastroianni Bnd, Falchero L, Chalabreysse L, Loire R, Ranche`re D, Souquet P-J, et al. Primary sarcomas of the lung: a clinicopathologic study of 12 cases. *Lung cancer (Amsterdam, Netherlands)*. 2002;38(3):283-9.
13. J J, J Z, W D. Primary pulmonary synovial sarcoma, a rare primary lung neoplasm: two case reports and review of the current literature. *Respirology (Carlton, Vic)*. 2008;13(5):748-50.
14. DB P, VI V, GT K, MA P, ED O, VI S, et al. Primary pulmonary sarcomas and carcinosarcomas—postoperative results and comparative survival analysis. *European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery*. 2003;23(4):461-6.

Tables

Table1. Clinical Characteristics of primary pulmonary sarcoma patients in the cohort.

| Characteristic | Case | No.(%) |
|-----------------------|-------------|---------------|
| Marriage | | |
| Unmarried | 108 | 43.7 |
| Married | 139 | 56.3 |
| Gender | | |
| Male | 142 | 57.5 |
| Female | 105 | 42.5 |
| Age(year) | | |
| <20 | 3 | 1.2 |
| 20~39 | 23 | 9.3 |
| 40~49 | 24 | 9.7 |
| 50~59 | 41 | 16.6 |
| 60~69 | 52 | 21.1 |
| 70~79 | 64 | 25.9 |
| ≥80 | 40 | 16.2 |
| Race | | |
| Black | 15 | 6.1 |
| White | 215 | 87.0 |
| Other | 17 | 6.9 |
| Site | | |
| Upper lobe | 132 | 53.4 |
| Middle lobe | 15 | 6.1 |
| Lower lobe | 90 | 36.4 |
| Main bronchus | 10 | 4.0 |
| Grade | | |
| Well | 11 | 4.5 |
| Moderate | 25 | 10.1 |
| Poor | 90 | 36.4 |
| Undifferentiated | 121 | 49.0 |

| Laterality | | |
|--------------------------|-----|------|
| Left | 114 | 46.2 |
| Right | 133 | 53.8 |
| Stage | | |
| Localized | 144 | 58.3 |
| Distant | 103 | 41.7 |
| Operation method | | |
| No surgery | 95 | 38.5 |
| Sublobectomy | 24 | 9.7 |
| Lobectomy | 104 | 42.1 |
| Total pneumonectomy | 24 | 9.7 |
| Operation | | |
| No | 95 | 38.5 |
| Yes | 152 | 61.5 |
| Radiotherapy | | |
| No | 211 | 85.4 |
| Yes | 36 | 14.6 |
| Chemotherapy | | |
| No | 170 | 68.8 |
| Yes | 77 | 31.2 |
| Neoplasm invasive | | |
| T1 | 102 | 41.3 |
| T2 | 49 | 19.8 |
| T3 | 38 | 15.4 |
| T4 | 35 | 14.2 |
| M1 | 23 | 9.3 |
| Lymph node status | | |
| N0 | 176 | 71.3 |
| N1 | 22 | 8.9 |

| | | |
|-----------------------|-----|------|
| N2 | 40 | 16.2 |
| N3 | 9 | 3.6 |
| Metastasis | | |
| No | 175 | 70.9 |
| Yes | 72 | 29.1 |
| First tumor | | |
| Yes | 165 | 66.8 |
| No | 82 | 33.2 |
| Tumor size(cm) | | |
| ≤3 | 47 | 19.0 |
| 3~5 | 54 | 21.9 |
| 5~7 | 43 | 17.4 |
| 7~10 | 54 | 21.9 |
| ≥10 | 49 | 19.8 |

Table2. Univariate and multivariate analysis for the training cohort.

| factors | Univariate Analysis | | Multivariate Analysis | |
|-----------------------|---------------------|----------|-----------------------|---------|
| | HR (95%CI) | P value | HR (95%CI) | P value |
| marriage | 0.553[0.355-0.862] | 0.009* | 0.492[0.307-0.787] | 0.003* |
| sex | 0.488[0.303-0.786] | 0.003* | 0.575[0.339-0.976] | 0.041* |
| age | 1.303[1.105-1.536] | 0.002* | 1.258[1.070-1.479] | 0.005* |
| race | 1.190[0.604-2.345] | 0.615 | - | - |
| site | 0.851[0.681-1.064] | 0.157 | - | - |
| grade | 1.546[1.163-2.057] | 0.003* | 1.601[1.148-2.233] | 0.006* |
| Laterality | 1.010[0.649-1.570] | 0.966 | - | - |
| stage | 1.910[1.225-2.977] | 0.004* | - | - |
| Operation method | 0.718[0.614-0.840] | <0.001** | - | - |
| operation | 0.333[0.209-0.529] | <0.001** | 0.462[0.280-0.762] | 0.002* |
| radiotherapy | 0.980[0.505-1.905] | 0.954 | - | - |
| chemotherapy | 1.166[0.723-1.880] | 0.528 | - | - |
| neoplasm invasiveness | 1.224[1.044-1.436] | 0.013* | - | - |
| Lymph node status | 1.438[1.114-1.857] | 0.005* | - | - |
| metastasis | 2.285[1.426-3.659] | 0.001* | 2.004[1.230-3.266] | 0.005* |
| First tumor | 0.932[0.569-1.524] | 0.778 | - | - |
| tumor size | 1.195[1.017-1.404] | 0.030* | 1.167[0.980-1.391] | 0.083 |

† HR=hazard ratio; CI = confidence interval.

Figures

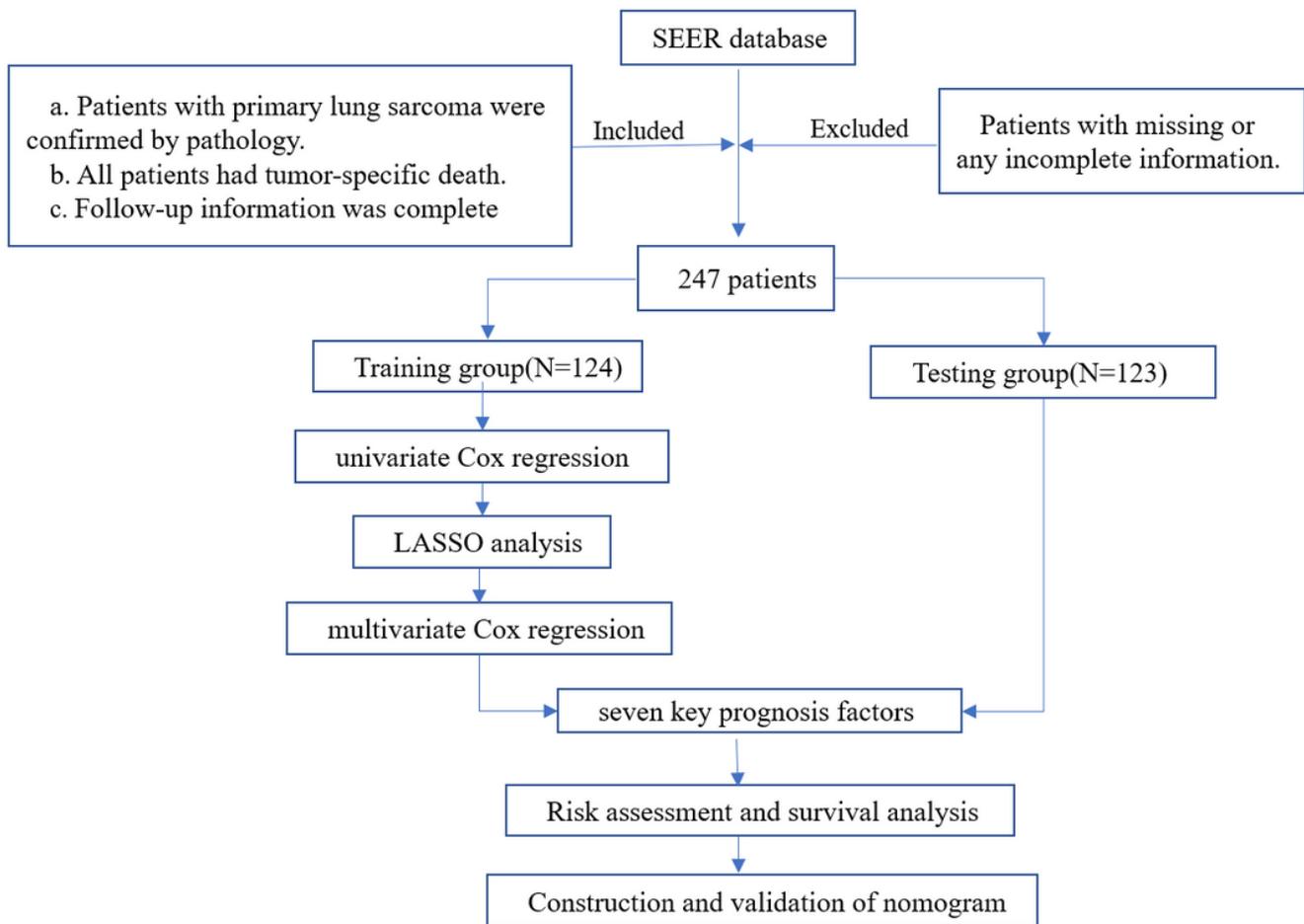


Figure 1

The workflow showing patient selection and analysis.

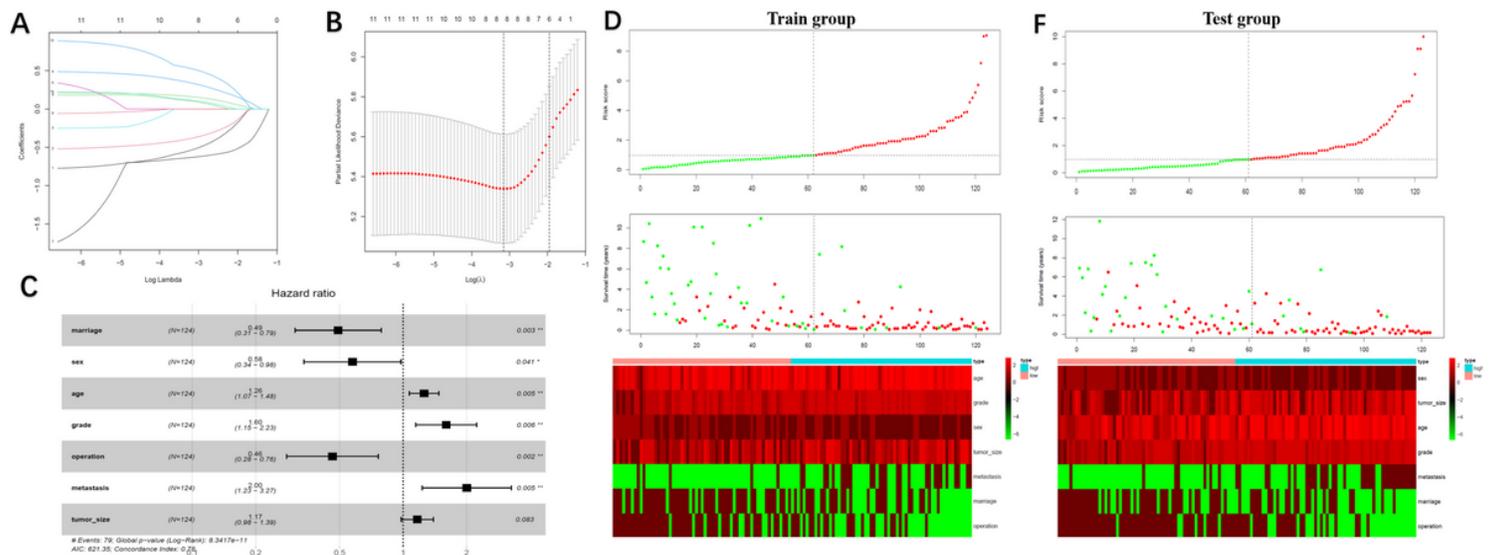


Figure 2

Identification of key prognostic factors in patients with primary pulmonary sarcoma. (A) LASSO coefficient profiles of the clinicopathological characteristics with PPS patients;(B) partial likelihood deviance was plotted corresponding $\log(\text{Lambda})$. LASSO, least absolute shrinkage and selection operator. (C) In the training group, seven significant prognostic factors associated with survival in patients with PPS were screened. Distribution of risk score, survival status of each patient, and heat map expression of seven prognosis factors in high-risk and low-risk groups in training set (D) and testing set (F).

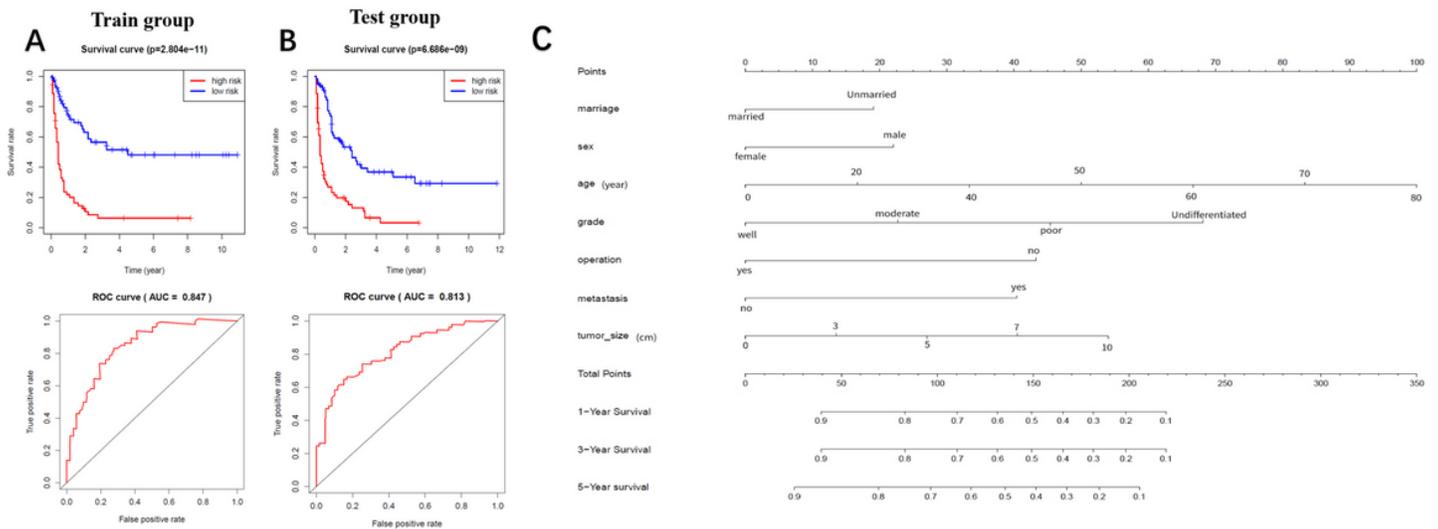


Figure 3

K-M survival curves along with the log-rank test and ROC analysis to evaluate performance of this risk score formula in training set (A) and testing set (B). AUC, area under the curve; K-M, Kaplan-Meier; ROC, receiver operating characteristic. (C) The nomogram for predicting probabilities of patients with 1-, 3- and 5-year overall survival.

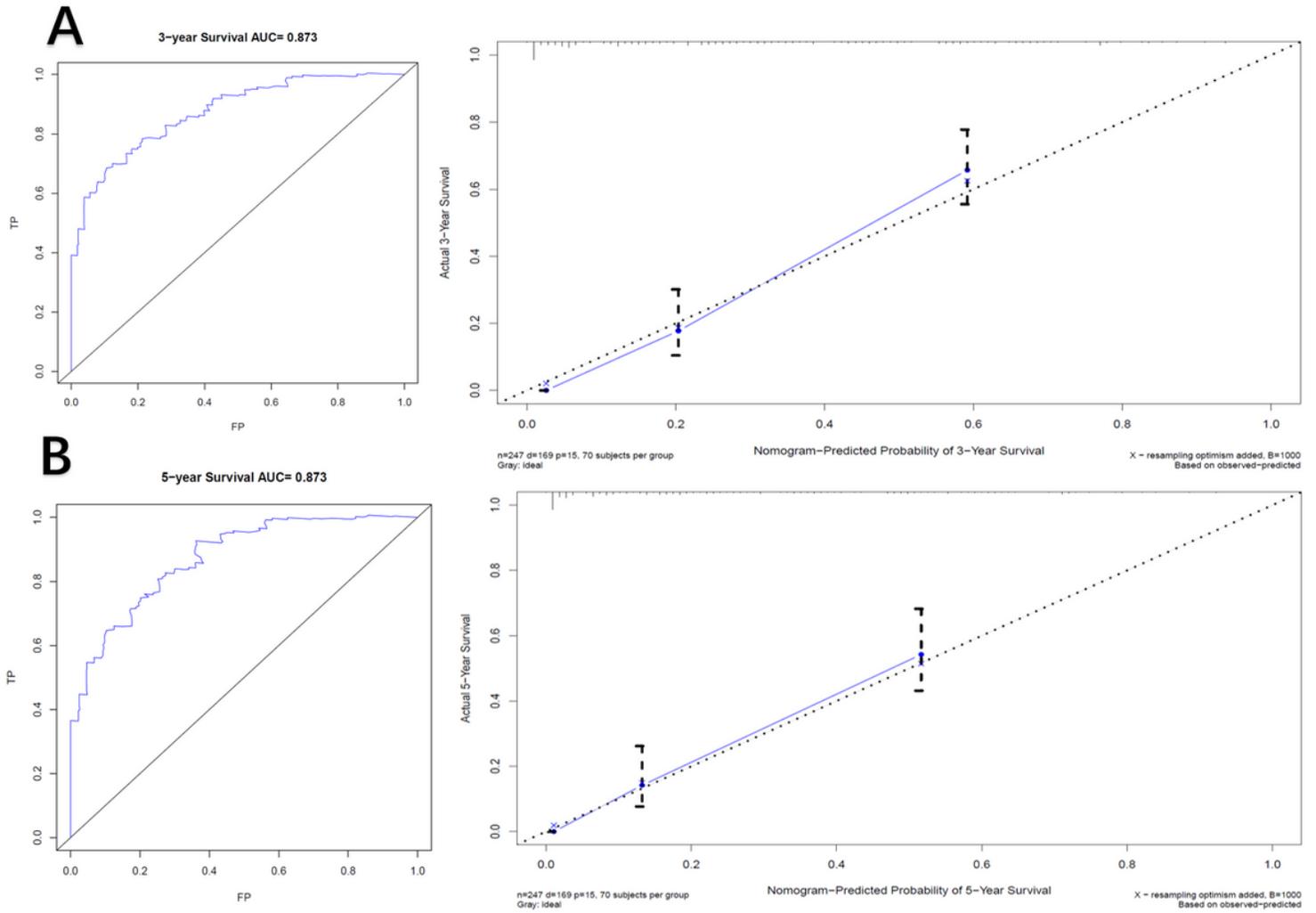


Figure 4

ROC curve based on the nomogram for overall survival probability. ROC, receiver operating characteristic; AUC, area under the curve. (A) 3-year Survival AUC was 0.873; (B) 5-year Survival AUC was 0.873; The calibration curve for predicting patient survival at (C) 3 years and (D) 5 years. Nomogram-predicted probability of overall survival is plotted on the x-axis; actual overall survival is plotted on the y-axis.