Development and Validation of a Nomogram to Predict Survival in Pulmonary Fibrosis Patients Admitted to the ICU

Journal:	*Science Progress*
Manuscript ID	SCI-25-1461
Manuscript Type:	Review
Date Submitted by the Author:	01-Jul-2025
Complete List of Authors:	Zhou, Yuanjun; Zhenjiang First People's Hospital, sun, ruxian; Zhenjiang First People's Hospital, Zhenjiang, China, Department of Critical Care Medicine tao, changyan; Zhenjiang First People's Hospital, Zhenjiang, China lv, jiao; Zhenjiang First People's Hospital, Zhenjiang, China, Department of Critical Care Medicine, Zhenjiang First People's Hospital
Keywords:	Nomogram, Pulmonary Fibrosis, Survival Rate, intensive care unit, Medical Information Mart for Intensive Care IV
Abstract:	ABSTRACT Background: Pulmonary interstitial fibrosis, a common end-stage manifestation of interstitial pneumonia, is a chronic and progressive lung disease marked by inflammatory cell infiltration and extensive fibrosis in the lung interstitium. Prompt identification of high-risk patients is crucial for guiding clinical decisions and determining the optimal timing for lung transplantation. This study aimed to develop and validate a nomogram to predict the overall survival rate of patients with pulmonary fibrosis (PF) in the intensive care unit (ICU). Methods: A total of 459 patients were included in this study from the Medical Information Mart for Intensive Care IV (MIMIC-IV) database. Within 24 hours of patient admission, 55 clinical indicators were collected. The least absolute selection and shrinkage operator (LASSO) was utilized with 10-fold cross-validation methods to select the optimal prognostic factors. Subsequently, multivariate COX regression analysis was performed to identify independent prognostic factors and construct the nomogram. Model performance and clinical utility were evaluated using the C-index, time-dependent receiver operating characteristic (ROC) curve, calibration curve, decision curve analysis (DCA), and Kaplan-Meier survival curve. Time-dependent ROC and DCA analyses were employed to compare the predictive value of the nomogram model with APSIII. Results: The 30-day, 90-day, and 180-day OS rates of 459 patients with PF were 68.41%, 62.75%, and 56.43%, respectively. Indicators included in the nomogram were age, temperature, RDW, PaO2, and APSIII. The C-index of the training set was 0.688 and that of the validation set was 0.678; the time-dependent ROC, calibration curves and DCA of the two groups showed good discrimination and accuracy. When compared to APSIII, the nomogram model demonstrated greater accuracy in predicting survival rates.

Development and Validation of a Nomogram to Predict Survival in Pulmonary Fibrosis Patients Admitted to the ICU

Yuanjun Zhou[1#], Changyan Tao[1#], Juan Xu[1], Jiao Lv[2], Ruxian Sun[1*]

[1]Department of Critical Care Medicine, Zhenjiang First People's Hospital, Zhenjiang, China.
[2]Department of Gastroenterology, Zhenjiang First People's Hospital, Zhenjiang, China.

[#]These authors contributed equally to this work.
[*]Correspondence to: Ruxian Sun
Ruxian Sun, Department of Critical Care Medicine, Zhenjiang First People's Hospital, No.8 Dian Li Road, Zhenjiang, Jiangsu 212000, China.
E-mail: drsunrx@163.com

ABSTRACT

Background: Pulmonary interstitial fibrosis, a common end-stage manifestation of interstitial pneumonia, is a chronic and progressive lung disease marked by inflammatory cell infiltration and extensive fibrosis in the lung interstitium. Prompt identification of high-risk patients is crucial for guiding clinical decisions and determining the optimal timing for lung transplantation. This study aimed to develop and validate a nomogram to predict the overall survival rate of patients with pulmonary fibrosis (PF) in the intensive care unit (ICU).

Methods: A total of 459 patients were included in this study from the Medical Information Mart for Intensive Care IV (MIMIC-IV) database. Within 24 hours of patient admission, 55 clinical indicators were collected. The least absolute selection and shrinkage operator (LASSO) was utilized with 10-fold cross-validation methods to select the optimal prognostic factors. Subsequently, multivariate COX regression analysis was performed to identify independent prognostic factors and construct the nomogram. Model performance and clinical utility were evaluated using the C-index, time-dependent receiver operating characteristic (ROC) curve, calibration curve, decision curve analysis (DCA), and Kaplan-Meier survival curve. Time-dependent ROC and DCA analyses were employed to compare the predictive value of the nomogram model with APSIII.

Results: The 30-day, 90-day, and 180-day OS rates of 459 patients with PF were 68.41%, 62.75%, and 56.43%, respectively. Indicators included in the nomogram were age, temperature, RDW, PaO_2, and APSIII. The C-index of the training set was 0.688 and that of the validation set was 0.678; the time-dependent ROC, calibration curves and DCA of the two groups showed good discrimination and accuracy. When compared to APSIII, the nomogram model demonstrated greater accuracy in predicting survival rates.

Conclusion: This study has successfully developed and validated the initial predictive model that integrates five clinical features. This model effectively forecasts short-term survival in ICU-admitted patients with PF for any reasons and promptly identifies high risk individuals.

Keywords: Nomogram, Pulmonary Fibrosis, Survival Rate

Introduction

Pulmonary interstitial fibrosis, a common end-stage feature of interstitial pneumonia, is a chronic and progressive lung disease characterized by inflammatory cell infiltration and extensive fibrosis in the lung interstitium. Clinical symptoms typically include cough, dyspnea, decreased activity tolerance, and resting hypoxemia in advanced stages. High-resolution computed tomography (CT) serves as a crucial diagnostic tool for this condition. There are various causes of interstitial pneumonia, each type presenting different pathophysiological characteristics, clinical manifestations, and prognoses. All forms of interstitial pneumonia have the potential to result in irreversible pulmonary fibrosis. Even after effective treatment or cure of the initial cause, pulmonary fibrosis can continue to progress [1]. Hence, it is not possible to predict the prognosis of pulmonary fibrosis solely based on its cause, as patients with the same underlying cause can exhibit varying clinical progressions and survival rates. Identifying high risk patients promptly is beneficial for guiding clinical decisions and determining the optimal timing for lung transplantation.

The clinical prediction model, being straightforward, convenient, and not constrained by subjective limitations, plays a key role in evaluating the severity of a patient's condition, garnering increasing attention in clinical practice. The combination of multiple parameters in the model collectively predicts the prognosis, resolving the issue where a single variable may not accurately predict the prognosis on its own. In 2012, Brett Ley et al. developed the Gender, Age, and Physiology (GAP) index model, incorporating gender, age, forced vital capacity (FVC%), and carbon monoxide diffusion capacity (DLCO%), to predict mortality in idiopathic pulmonary fibrosis. The GAP index categorized patients into three stages (Stage I, Stage II, and Stage III), corresponding to 1-year mortality rates of 6%, 16%, and 39%, respectively, results consistent with subsequent follow-up assessments [2].

In this study, we reviewed clinical data from Medical Information Mart for Intensive Care IV (MIMIC-IV) patients with pulmonary fibrosis to establish a prognostic scoring system and nomogram for predicting survival in pulmonary fibrosis. The goal is to enhance the capacity for predicting the prognosis of pulmonary fibrosis patients and for guiding clinical decision-making.

Methods

Data source

This study employed a retrospective observational design, utilizing data from the publicly accessible MIMIC-IV version 2.2 critical care dataset, specifically focusing on patients treated at the Beth Israel Deaconess Medical Center from 2008 to 2019. To ensure compliance with regulatory standards, researcher Ruxian Sun (Record ID: 61,773,273) obtained a Collaborative Institutional training Initiative (CITI) certification and the necessary permissions for accessing the MIMIC-IV database. The study meticulously adhered to the STROCSS guidelines. As a retrospective analysis using the publicly available MIMIC database, this study does not involve prospective clinical trials or require a Clinical Trial Number.

Study population

Patients admitted to the ICU at Beth Israel Deaconess Medical Center for any reasons and with a final primary diagnosis of pulmonary fibrosis according to the 9th and 10th Revisions of the International Classification of Diseases were included in this study. For those with multiple admissions, only data from their initial Intensive Care Unit (ICU) admission were analyzed. The exclusion criteria were: (1) no diagnosis of pulmonary fibrosis, (2) age under 18, and (3) missing survival data. Ultimately, 459 patients were enrolled in the study (Figure 1).

Feature extraction

Data were collected using Structured Query Language (SQL) with PostgreSQL version 14.2 to extract baseline patient characteristics. The collected data encompassed demographics, vital signs, laboratory tests, comorbidities, and scoring systems. Demographics included age, gender, weight, and smoking history. Vital signs were comprised of heart rate, mean arterial blood pressure (MAP), respiratory rate (RR), temperature, and 24-hour urine volume. Laboratory tests included red blood cell (RBC) count, white blood cell (WBC) count, hemoglobin, hematocrit, red cell distribution width (RDW), platelet count, creatinine, blood urea nitrogen (BUN), potassium, sodium, chloride, calcium, blood glucose, pH, partial pressure of arterial oxygen (PaO_2), carbon dioxide partial pressure ($PaCO_2$), lactic acid, bicarbonate, activated partial thromboplastin time (APTT), prothrombin time (PT),

international normalized ratio (INR). Scoring systems included systemic inflammatory response syndrome (SIRS), acute physiology score III (APSIII), sequential organ failure assessment (SOFA), Glasgow coma scale (GCS), simplified acute physiology score II (SAPSII), Oxford acute severity of illness score (OASIS), and logistic organ dysfunction system (LODS). Comorbidities included myocardial infarction, congestive heart failure, peripheral vascular disease, cerebrovascular disease, chronic obstructive pulmonary disease (COPD), rheumatic disease, renal disease, malignant cancer, atrial fibrillation, hypertension, asthma, pulmonary arterial hypertension, obstructive sleep apnea syndrome (OSAS), gastroesophageal reflux, osteoporosis, diabetes, liver disease, and the Charlson comorbidity index. Variables with more than 20% missing data, such as height, albumin, FiO_2 and respiratory support methods, were excluded. All variables mentioned above were collected within 24 hours of patients' admission to the ICU. If a patient had multiple measurements within 24 hours of ICU admission, data from the initial measurement were used.

Statistical Analysis

Continuous variables were expressed as mean ± standard deviation (SD) or median and interquartile range (IQR) according to their distribution, while categorical variables were presented as proportions. The normality of continuous parameters was assessed using the Kolmogorov-Smirnov test. Normally distributed continuous variables were analyzed using the t-test or ANOVA, whereas non-normally distributed variables were assessed using the Mann-Whitney U test or Kruskal-Wallis test. To handle missing data, we employed multiple imputation using a random forest algorithm, implemented through the 'mice' package in R software, where the algorithm was trained using other non-missing variables [3, 4].

The primary endpoint was overall survival (OS), which was defined as the time from ICU admission until death. All patients were randomly divided into the training and validation sets, with a ratio of 7:3. The Least Absolute Shrinkage and Selection Operator (LASSO) Cox regression method was used for selecting the appropriate predictive features (features with non-zero coefficients) based on the lambda-min value in the training set, then followed by a multivariate analysis using Cox's proportional hazards regression model to identify independent prognostic risk factors from these clinical variables [5–7]. Nomograms were constructed based on independent prognostic factors identified in the training set [8]. To ensure no collinearity among the selected covariates, we calculated the variance inflation factor (VIF), defining collinearity as VIF > 2.0. We also confirmed that the nomogram adhered to the "10 EPV" guideline, which requires at least ten positive samples for each predictive variable. The model's discrimination ability was evaluated using the concordance index (C-index) and the area under the time-dependent receiver operating characteristic (ROC) curve. Calibration curves assessed the consistency between observed and predicted outcomes, and decision curve analysis (DCA) evaluated the net clinical benefit [9]. Using the nomogram, we calculated total points for each patient and employed the restricted cubic spline (RCS) regression model to analyze the relationship between predicted points and prognosis. Patients were then stratified into low and high risk groups based on the median of predicted points. The survival outcomes of these groups were analyzed using Kaplan-Meier survival curves and evaluated by the log-rank test. Additionally, we compared the predictive model with the prognostic value of APSIII using ROC and DCA curves. All analyses were conducted using R software (version 4.3.3), with a two-tailed p-value < 0.05 considered statistically significant.

Results

Patient characteristics

In total, 459 patients were enrolled in this study, 214 (46.62%) were female. Their median age was 76 (IQR: 66-83) years. The 30-day, 90-day, and 180-day OS rates were 68.41%, 62.75%, and 56.43%, respectively. The 30-day, 90-day, and 180-day OS rates in the training set were 71.01%, 63.77%, and 56.52%, respectively, contrast to those of 67.29%, 62.31%, and 56.39% in the validation set, respectively. Baseline characteristics showed no significant differences between the training and validation sets. For a comprehensive overview of all patients' baseline characteristics, please refer to Table 1.

Characteristics selection and development of the nomogram

From a pool of fifty-five variables, only nine were retained in the LASSO logistic regression model, selected based on the binomial deviance minimum criteria (Figure 2). Subsequently, these nine variables were incorporated into the multivariate Cox regression model, resulting in the identification of five variables with a p-value <0.05. These variables were deemed independent risk prognostic factors for OS in patients with pulmonary fibrosis. The final set of five variables included in the multivariate Cox regression analysis were: age (HR: 1.0122; 95% CI: 1.0015-1.0246), temperature (HR: 0.7948; 95% CI: 0.6582-0.9599), RDW (HR: 1.0829; 95% CI: 1.0101-1.1610), PaO_2 (HR: 0.9965; 95% CI: 0.9950-0.9981), and APSIII (HR: 1.0106; 95% CI: 1.0003-

1.0209) (Table 2). Utilizing this model, we developed a nomogram to predict OS in ICU patients with pulmonary fibrosis (Figure 3).

Evaluation and validation of the nomogram

The VIF values for the mentioned variables were all below 2.0, indicating no collinearity in our model. To assess the discriminative ability of the nomogram, we employed the C-index and time-dependent ROC analysis. In the training set, the C-index was 0.688 (95% CI: 0.651-0.725), while in the validation set, it was 0.678 (95% CI: 0.647-0.709). Regarding 30-day OS, the area under the curve (AUC) was 0.762 (95% CI: 0.702-0.814) in the training set and 0.741 (95% CI: 0.696-0.788) in the validation set. Similarly, for 90-day OS, the AUC in the training set was 0.750 (95% CI: 0.696-0.802), and in the validation set, it was 0.731 (95% CI: 0.686-0.776). As for 180-day OS, the AUC was 0.752 (95% CI: 0.701-0.804) in the training set and 0.739 (95% CI: 0.695-0.781) in the validation set (Figure 4). Additionally, calibration curves were employed to compare predicted and actual survival rates. These curves demonstrated good agreement between nomogram predictions and observed survival rates for 30-day, 90-day, and 180-day OS in both the training and validation sets (Figure 5). DCA was utilized to evaluate the clinical utility of the decision-making process [10]. Figure 6 illustrates the DCA for the nomogram, highlighting its superior predictive performance compared to scenarios where all patients are considered dead or none are considered dead [9, 10]. Additionally, DCA indicated a significant net clinical benefit associated with the nomogram prediction model. In summary, the nomogram model demonstrated robust predictive accuracy and clinical relevance for patients diagnosed with pulmonary fibrosis.

Risk stratification based on the nomogram

Figure 7 displays a RCS regression model, indicating a linear relationship (P for nonlinear = 0.691) between the predictive points derived from the nomogram and the prognosis. Consequently, higher scores correspond to decreased OS rate. According to the median value (112.07), individuals in both the training and validation sets were subsequently categorized into low risk and high risk groups. The Kaplan-Meier curves for both the training and validation sets revealed significant differences in survival between low risk and high risk patients at 30 days, 90 days, and 180 days (Figure 8).

Comparison of predictive value between the nomogram model and APSIII

The nomogram prediction model was compared with the APSIII (Figure 9). In the training set, the AUC of the nomogram prediction model for 30-day, 90-day, and 180-day OS was 0.762 (95% CI: 0.702-0.814), 0.750 (95% CI: 0.696-0.802), and 0.752 (95% CI: 0.701-0.804), respectively. In comparison, the AUC of APSIII for 30-day, 90-day, and 180-day OS was only 0.758 (95% CI: 0.699-0.817), 0.713 (95% CI: 0.654-0.774), and 0.699 (95% CI: 0.638-0.755), respectively. This indicated that the nomogram was more effective than the single APSIII scoring system in predicting OS. The DCA curves showed that our model provided a superior net clinical benefit compared to the APSIII scoring system for 30-day, 90-day, and 180-day predictions.

Table 1. Demographic and clinical characteristics of the training and validation sets

Variables	Total (n = 459)	Training set (n = 321)	Validation set (n = 138)	P value
Demographics				
Age, years	76 (66-83)	76 (67-82)	76 (66-84)	0.754
Female, n (%)	214 (46.62%)	142 (44.24%)	72 (52.17%)	0.118
Weight, kg	34.4 (28.9-40.0)	34.2 (28.6-39.7)	34.6 (29.3-40.4)	0.554
Smoker, n (%)	105 (22.88%)	74 (23.05%)	31 (22.46%)	0.890
Vital signs				
HR, beats/min	89 (76-103)	88 (76-102)	89 (76-104)	0.321
MAP, mmHg	80 (71-92)	80 (71-92)	79 (71-91)	0.909
RR, times/min	20 (16-26)	20 (16-26)	20 (17-26)	0.593
Temperature, ℃	36.7 (36.4-36.9)	36.7 (36.4-36.9)	36.7 (36.3-36.9)	0.731
Laboratory tests				
RBC, m/uL	3.71±0.76	3.72±0.73	3.70±0.81	0.806
WBC, K/uL	11.7 (8.5-15.7)	11.2 (8.3-15.2)	11.9 (8.8-16.0)	0.333
Hemoglobin, g/dL	11.12±2.22	11.13±2.14	11.09±2.43	0.852
Hematocrit, %	33.99±6.59	34.09±6.34	33.74±7.14	0.600
RDW, %	14.8 (13.7-16.1)	14.7 (13.7-16.1)	14.8 (13.7-16.1)	0.844
Platelet, K/uL	227 (161-307)	227 (159-302)	227 (172-316)	0.446
Creatinine, mg/dL	1.0 (0.8-1.4)	1.0 (0.8-1.4)	1.0 (0.8-1.3)	0.856
BUN, mg/dL	21 (15-32)	21 (15-32)	22 (15-31)	0.471
Potassium, mEq/L	4.2 (3.8-4.6)	4.2 (3.8-4.6)	4.2 (3.9-4.6)	0.553
Sodium, mEq/L	138 (135-140)	138 (135-141)	138 (135-140)	0.368
Chloride, mEq/L	101 (97-106)	101 (97-106)	101 (98-105)	0.837
Calcium, mg/dL	8.5 (8.1-9.0)	8.5 (8.1-9.0)	8.5 (8.0-8.9)	0.665
Blood glucose, mg/dL	128 (105-169)	127 (104-166)	134 (107-173)	0.109
PH	7.40 (7.34-7.45)	7.40 (7.33-7.46)	7.40 (7.35-7.44)	0.915
PaO_2, mmHg	86 (52-161)	88 (53-160)	80 (51-172)	0.505
$PaCO_2$, mmHg	42 (36-49)	43 (36-49)	41 (37-46)	0.573
Lactic acid, mmol/L	1.7 (1.2-2.5)	1.7 (1.2-2.5)	1.7 (1.2-2.5)	0.885
Hydrocarbonate, mEq/L	24 (22-27)	24 (22-27)	24 (22-27)	0.677
APTT, s	30.0 (26.8-35.9)	30.3 (26.9-35.9)	29.6 (26.4-35.7)	0.569
PT, s	13.6 (12.4-16.2)	13.6 (12.5-16.3)	13.6 (12.2-15.9)	0.588
INR	1.2 (1.1-1.5)	1.2 (1.1-1.5)	1.2 (1.1-1.4)	0.477
24-hour urine volume, ml	1400 (900-2265)	1395 (875-2170)	1449 (1001-2571)	0.157
Clinical score				
SIRS	3 (2-3)	3 (2-3)	3 (2-3)	0.940
APSIII	44 (34-58)	43 (34-58)	45 (34-57)	0.790
SOFA	4 (2-6)	4 (2-6)	4 (2-6)	0.201
GCS	14 (13-15)	14 (13-15)	14 (13-15)	0.473
SAPSII	35 (28-42)	35 (28-43)	35 (28-41)	0.990
OASIS	32 (27-38)	32 (27-38)	32 (26-37)	0.246
LODS	4 (2-6)	4 (2-7)	4 (2-6)	0.589
Comorbidities, n (%)				
Myocardial infarct	78 (16.99%)	56 (17.45%)	22 (15.94%)	0.694
Congestive heart failure	209 (45.53%)	146 (45.48%)	63 (45.65%)	0.973
Peripheral vascular disease	65 (14.16%)	47 (14.64%)	18 (13.04%)	0.652
Cerebrovascular disease	48 (10.46%)	37 (11.53%)	11 (7.97%)	0.254
COPD	237 (51.63%)	170 (52.96%)	67 (48.55%)	0.386

Rheumatic disease	47 (10.24%)	33 (10.28%)	14 (10.14%)	0.965
Renal disease	109 (23.75%)	79 (24.61%)	30 (21.74%)	0.507
Malignant cancer	56 (12.20%)	39 (12.15%)	17 (12.32%)	0.959
Atrial fibrillation	153 (33.33%)	112 (34.89%)	41 (29.71%)	0.280
Hypertension	311 (67.76%)	217 (67.60%)	94 (68.12%)	0.914
Asthma	48 (10.46%)	36 (11.21%)	12 (8.70%)	0.419
Pulmonary arterial hypertension	25 (5.45%)	17 (5.30%)	8 (5.80%)	0.828
OSAS	55 (11.98%)	38 (11.84%)	17 (12.32%)	0.884
Esophageal reflux	129 (28.10%)	89 (27.73%)	40 (28.99%)	0.783
Osteoporosis	39 (8.50%)	30 (9.35%)	9 (6.52%)	0.320
Diabetes	136 (29.63%)	101 (31.46%)	35 (25.36%)	0.189
Liver disease	40 (8.71%)	30 (9.35%)	10 (7.25%)	0.465
Charlson comorbidity index	6 (5-8)	6 (5-8)	6 (5-8)	0.209
Outcome				
30-day OS rate, n (%)	314 (68.41%)	98 (71.01%)	216 (67.29%)	0.431
90-day OS rate, n (%)	288 (62.75%)	88 (63.77%)	200 (62.31%)	0.766
180-day OS rate, n (%)	259 (56.43%)	78 (56.52%)	181 (56.39%)	0.979

Continuous variables are presented as mean ± SD if normally distributed, and median (interquartile range) if not normally distributed. Categorical variables are presented as number of patients (%). HR: heart rate; MAP: mean arterial blood pressure; RR: respiratory rate; RBC: red blood cell; WBC: white blood cell; RDW: red cell distribution width; BUN: blood urea nitrogen; PH: potential of hydrogen; PaO_2: partial pressure of arterial oxygen; $PaCO_2$: carbon dioxide partial pressure; APTT: activated partial thromboplastin time; PT: prothrombin time; INR: international normalized ratio; SIRS: systemic inflammatory response syndrome; APSIII: acute physiology score III; SOFA: sequential organ failure assessment; GCS: Glasgow coma scale; SAPSII: simplified acute physiology score II; OASIS: oxford acute severity of illness score; LODS: logistic organ dysfunction system; COPD: chronic obstructive pulmonary disease; OSAS: obstructive sleep apnea syndrome; OS: overall survival.

Table 2. Multivariate cox regression analysis based on LASSO regression results

Variables	Coefficients	HR (95%CI)	P value
Age	0.012156	1.0122 (1.0015-1.0246)	**0.0493**
Temperature	-0.229618	0.7948 (0.6582-0.9599)	**0.0171**
WBC	0.023057	1.0233 (0.9990-1.0483)	0.0605
RDW	0.079686	1.0829 (1.0101-1.1610)	**0.0249**
PaO_2	-0.003471	0.9965 (0.9950-0.9981)	**<0.0001**
SIRS	0.151512	1.1636 (0.9786-1.3836)	0.0864
APSIII	0.010508	1.0106 (1.0003-1.0209)	**0.0434**
OASIS	0.013335	1.0134 (0.9907-1.0367)	0.2495
LODS	0.055595	1.0572 (0.9789-1.1417)	0.1566

LASSO: least absolute shrinkage and selection operator; HR: hazard ratio; CI: confidence interval; WBC: white blood cell; RDW: red cell distribution width; PaO_2: partial pressure of arterial oxygen; SIRS: systemic inflammatory response syndrome; APSIII: acute physiology score III; OASIS: oxford acute severity of illness score; LODS: logistic organ dysfunction system.

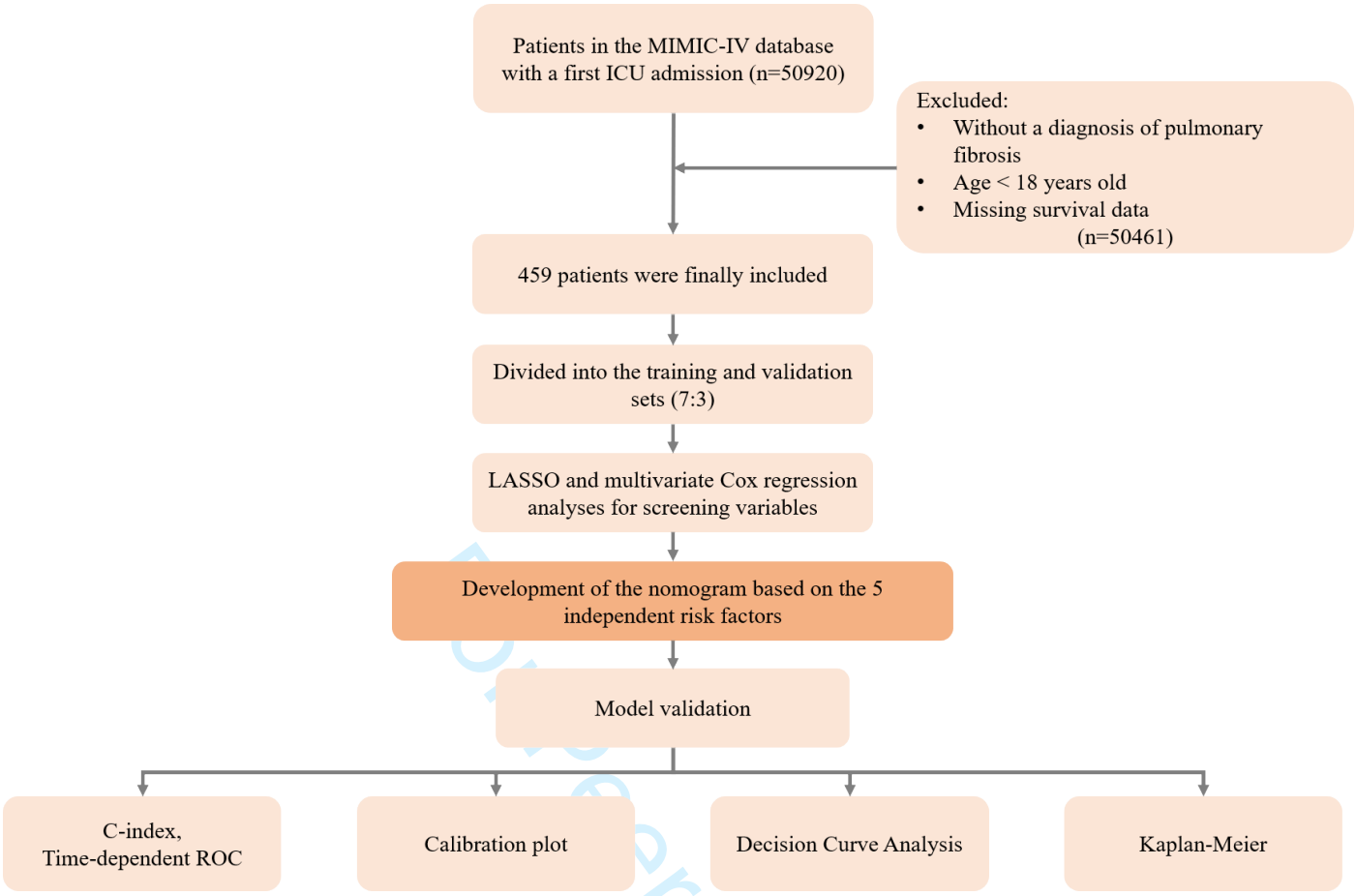


Figure 1. Workflow of the study. MIMIC-IV: Medical Information Mart for Intensive Care IV; ICU: intensive care unit; LASSO: least absolute shrinkage and selection operator. ROC: receiver operating characteristic.

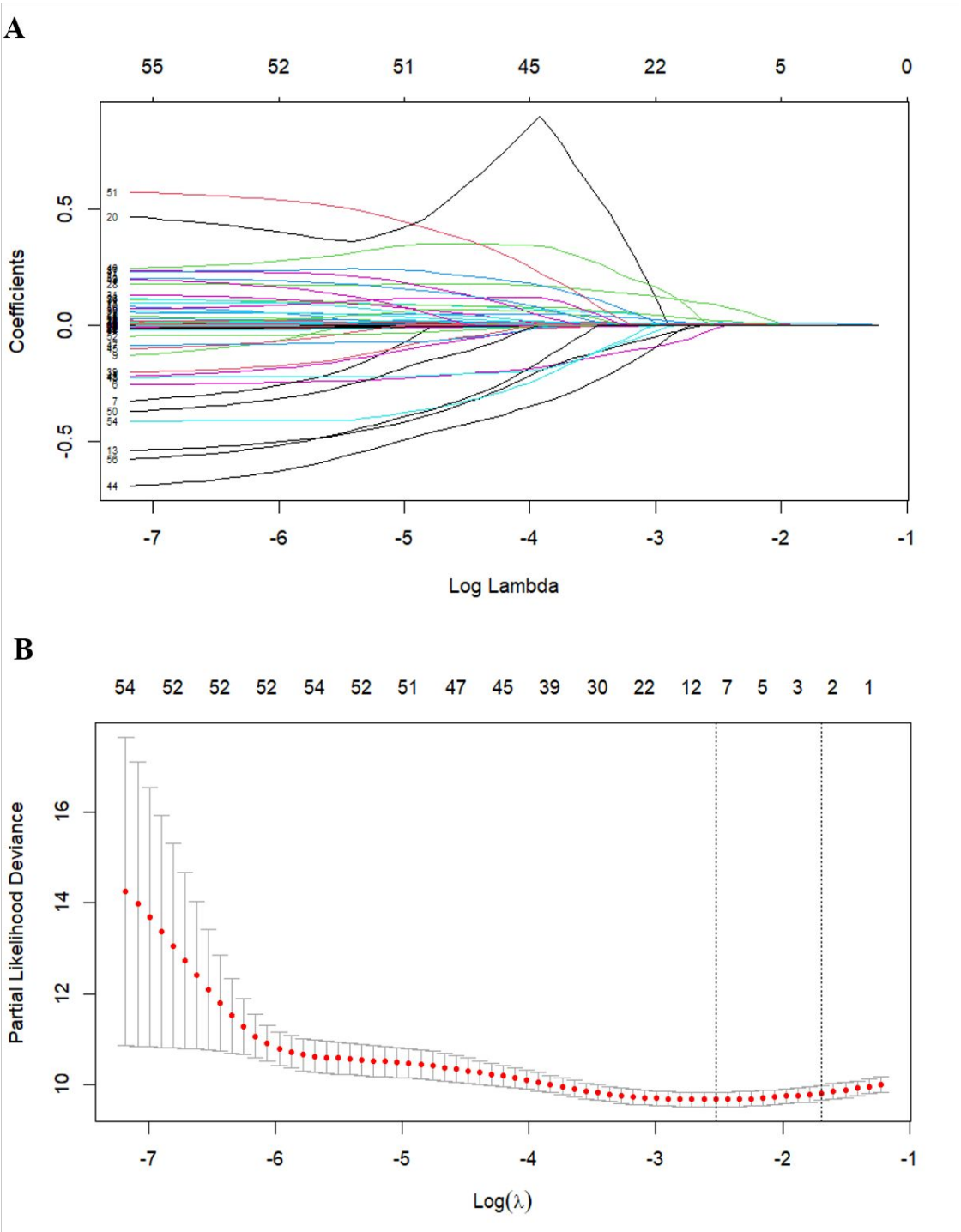


Figure 2. The LASSO regression analysis to select the potential variables. A total of fifty-five variables were initially included and nine variables were finally selected for further analysis. The LASSO coefficient analysis of the clinical features **(A)**. Cross-validation was conducted using a 10-fold approach to optimize the tuning parameter selection in the LASSO model **(B)**. LASSO: least absolute shrinkage and selection operator.

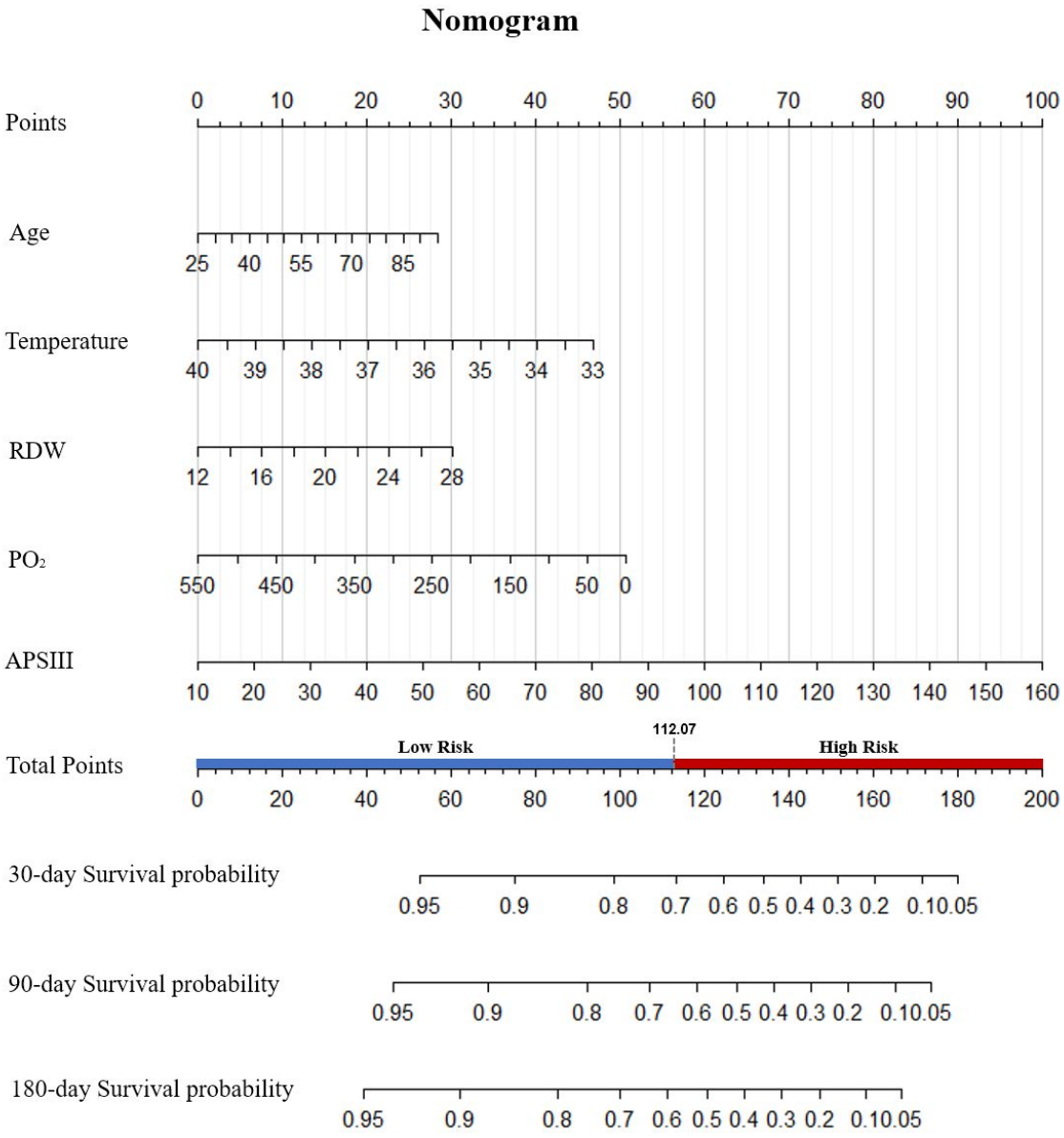


Figure 3. The nomogram for predicting 30-day, 90-day, and 180-day survival probabilities in patients with pulmonary fibrosis.

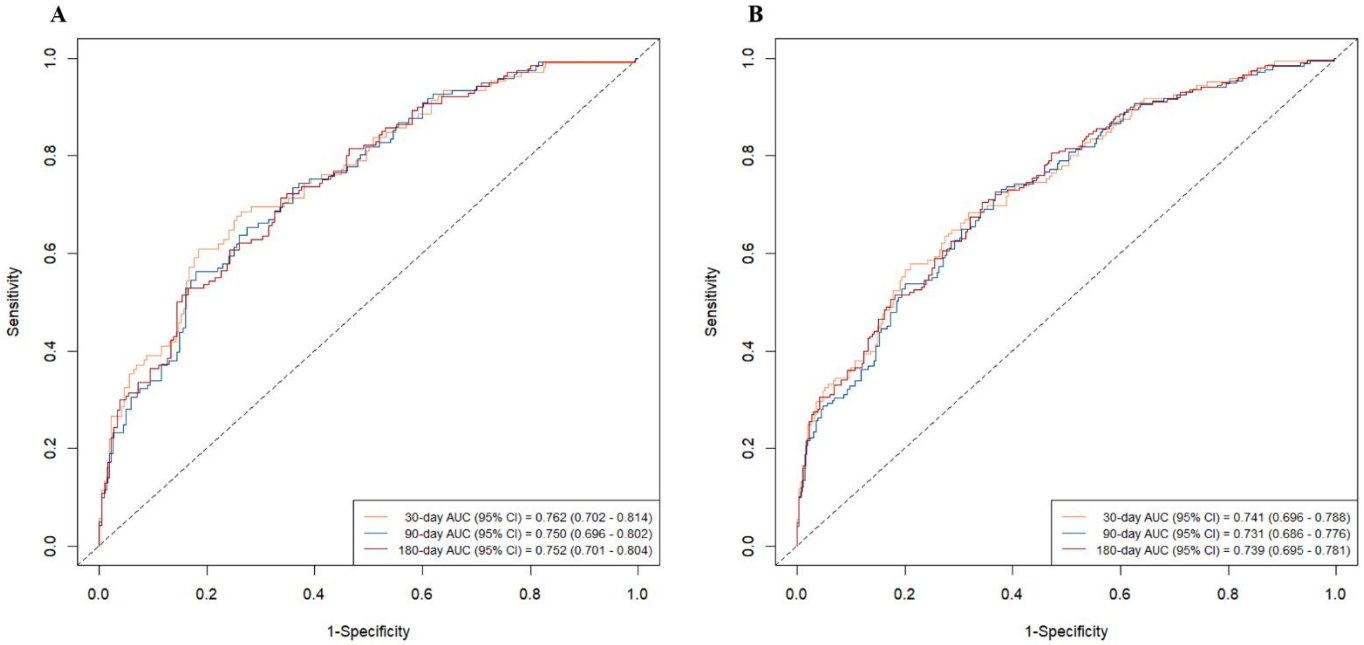


Figure 4. Time-dependent ROC curves using the nomogram predict overall survival probabilities within 30-day, 90-day, and 180-day in the training set **(A)** and validation set **(B)**. AUC: area under the curve; CI: confidence interval; ROC: receiver operating characteristic.

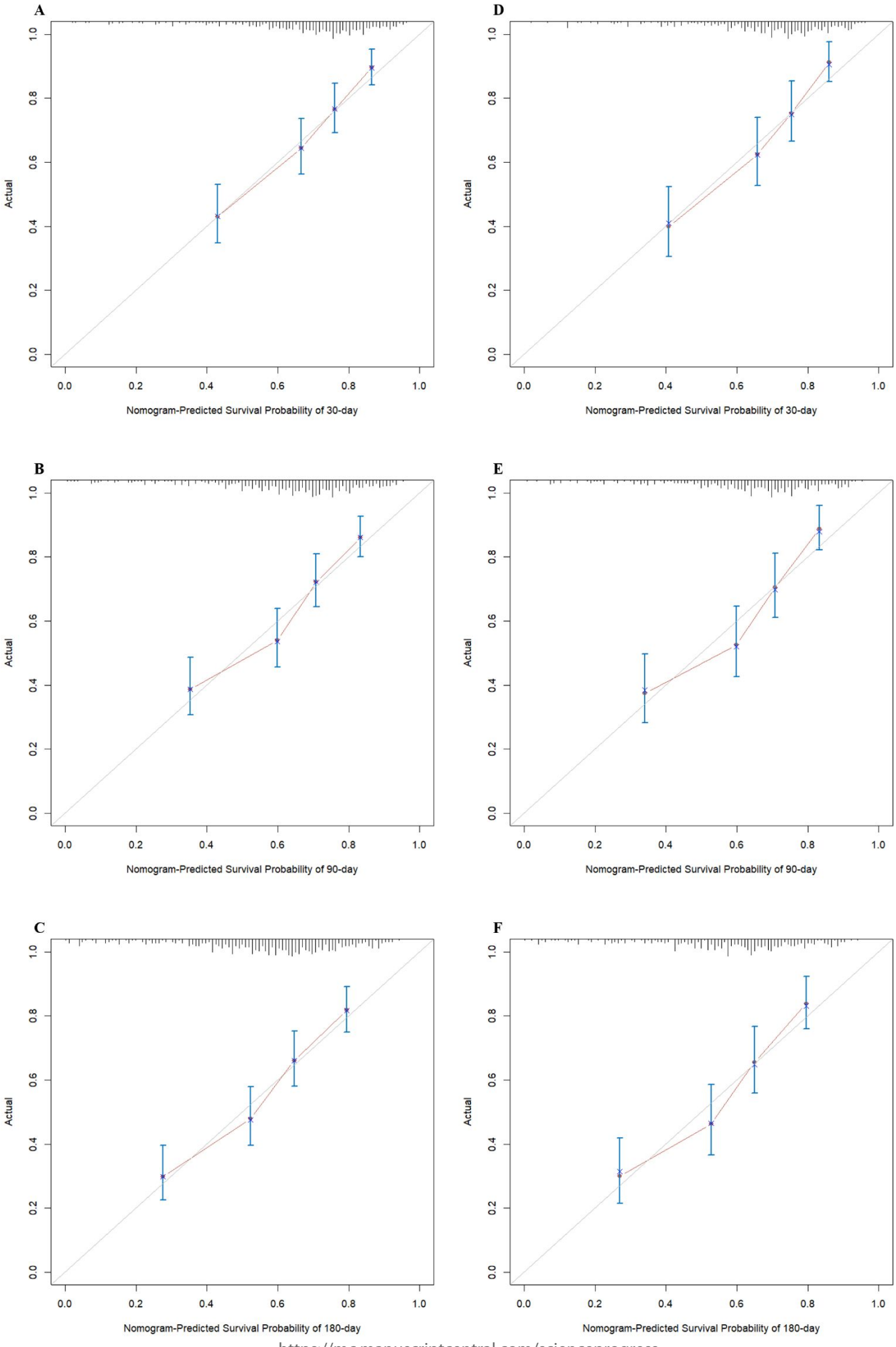


Figure 5. Calibration curves for OS at 30-day, 90-day, and 180-day in patients with pulmonary fibrosis. 30-day OS rate in the training set **(A)**; 90-day OS rate in the

training set **(B)**; 180-day OS rate in the training set **(C)**; 30-day OS rate in the validation set **(D)**; 90-day OS rate in the validation set **(E)**; 180-day OS rate in the validation set **(F)**. The horizontal axis depicts the survival rate predicted by the nomogram, while the vertical axis represents the actual survival rate. The grey diagonal line indicates that the predicted survival rate perfectly matches the actual survival rate. OS: overall survival.

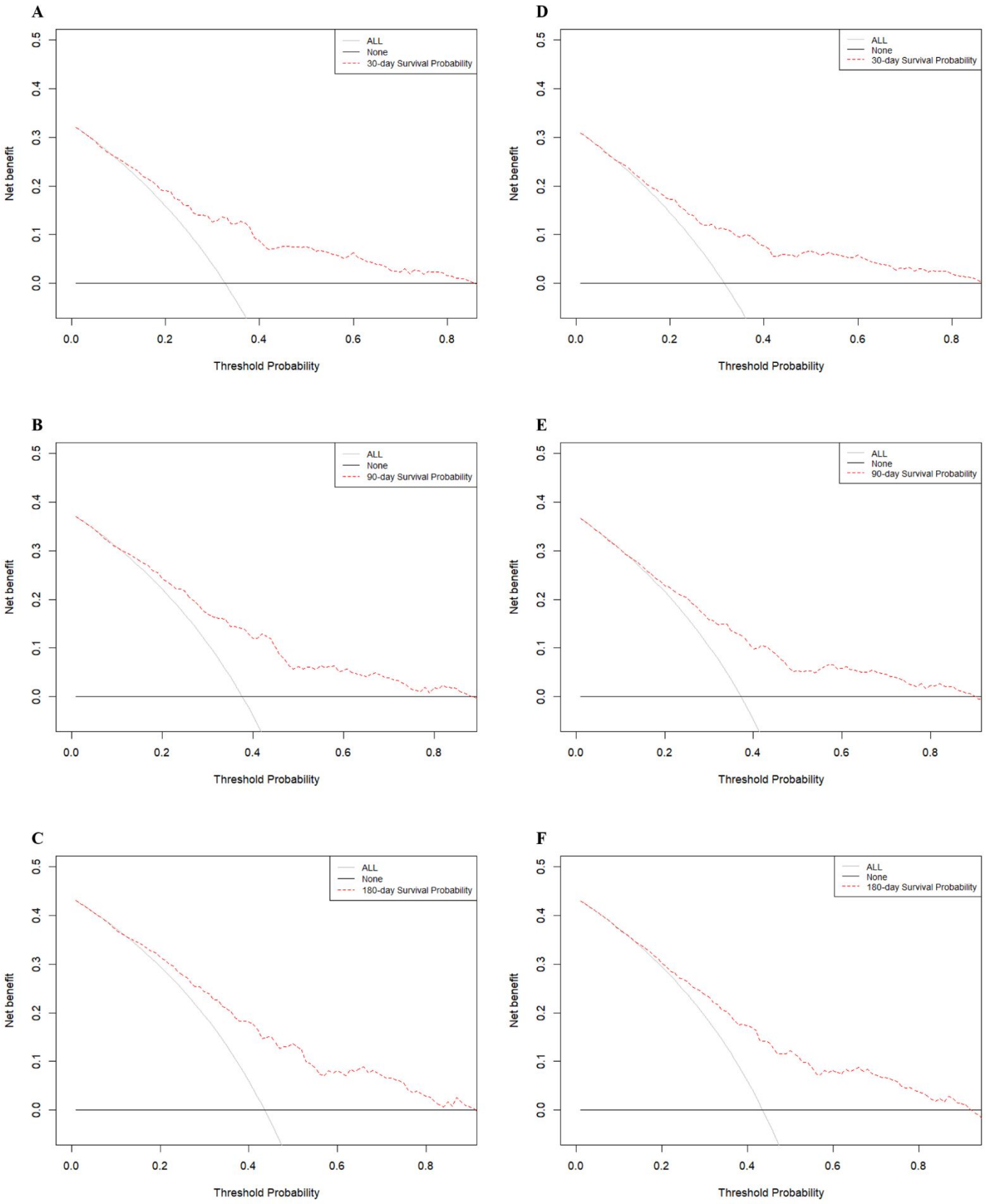


Figure 6. Decision curve analysis of the training set for 30-day **(A)**, 90-day **(B)**, and 180-day **(C)** survival probabilities. Decision curve analysis of the validation set for 30-day **(D)**, 90-day **(E)**, and 180-day **(F)** survival probabilities. The y-axis measures the net benefit. The gray line represents the assumption that all patients are deceased. The black line represents the assumption of no patient deaths. The red dotted line represents the nomogram.

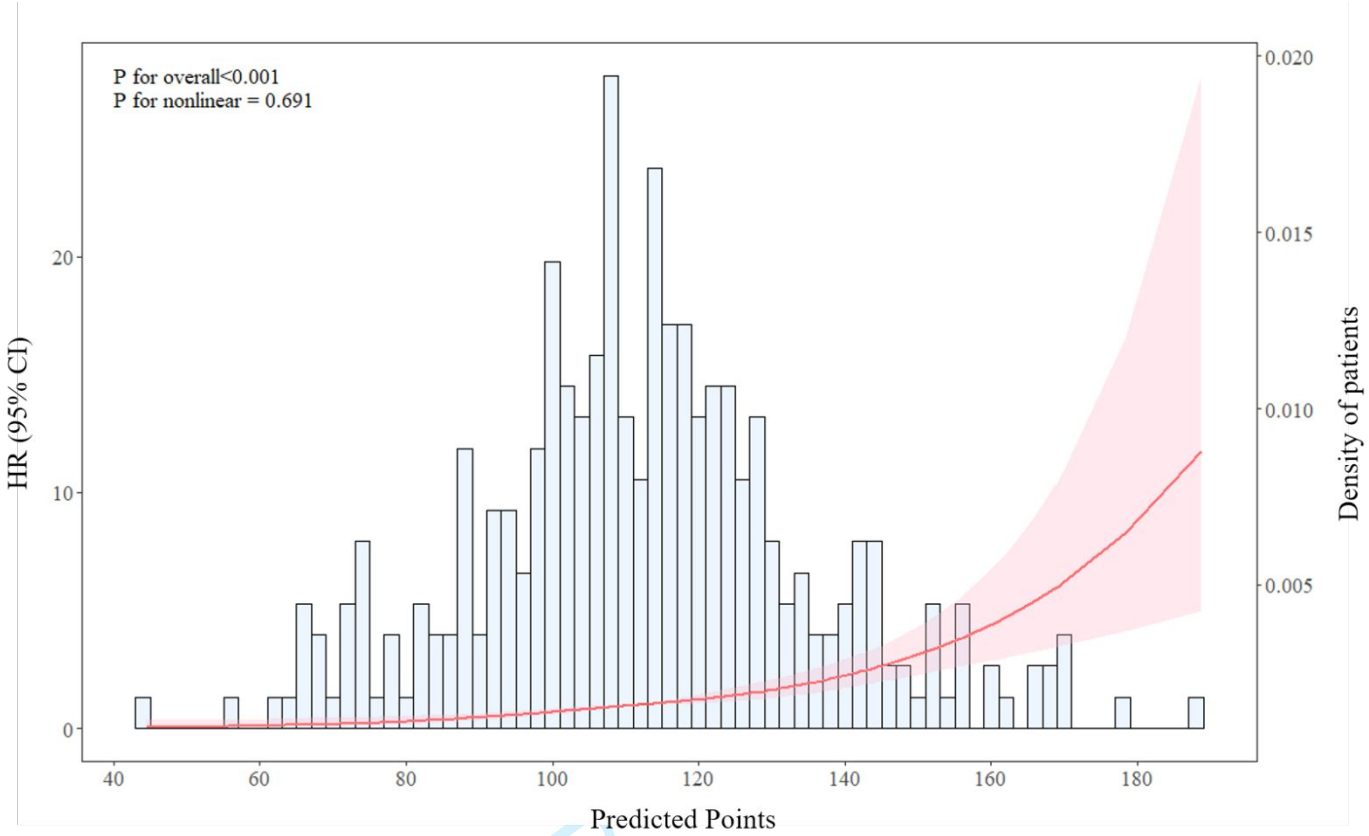


Figure 7. Restricted cubic spline regression analysis for the predicted points derived from the nomogram. The heavy central lines represent the estimated hazard ratios, with shaded ribbons indicating the corresponding 95% confidence intervals. The histogram illustrates the density of patients. HR: hazard ratio; CI: confidence interval.

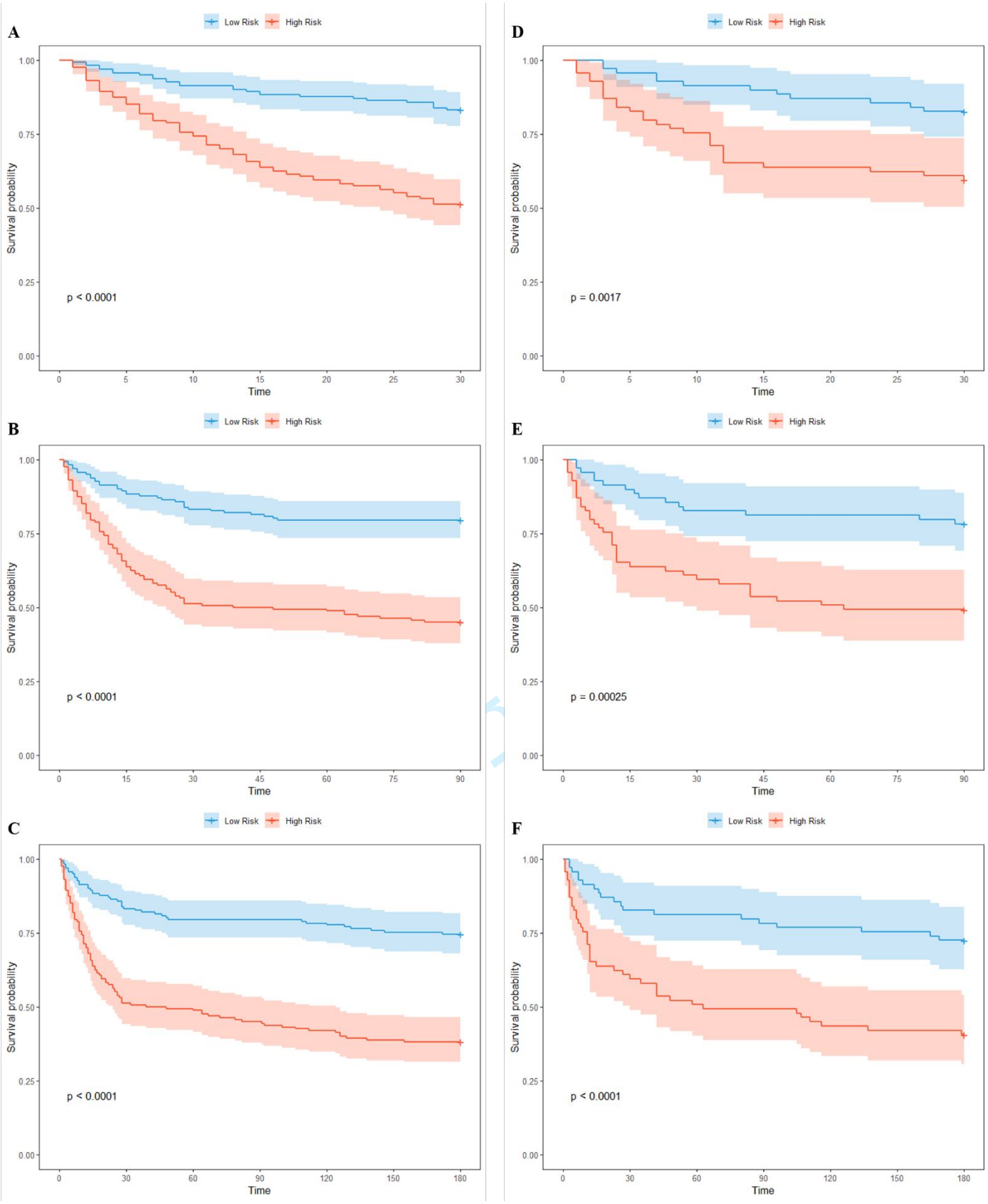


Figure 8. Kaplan-Meier survival curves of OS for patients with pulmonary fibrosis at 30-day **(A)**, 90-day **(B)**, and 180-day **(C)** in the training set, and at 30-day **(D)**, 90-day **(E)**, and 180-day **(F)** in the validation set. Patients with low risk were associated with better survival. OS: overall survival.

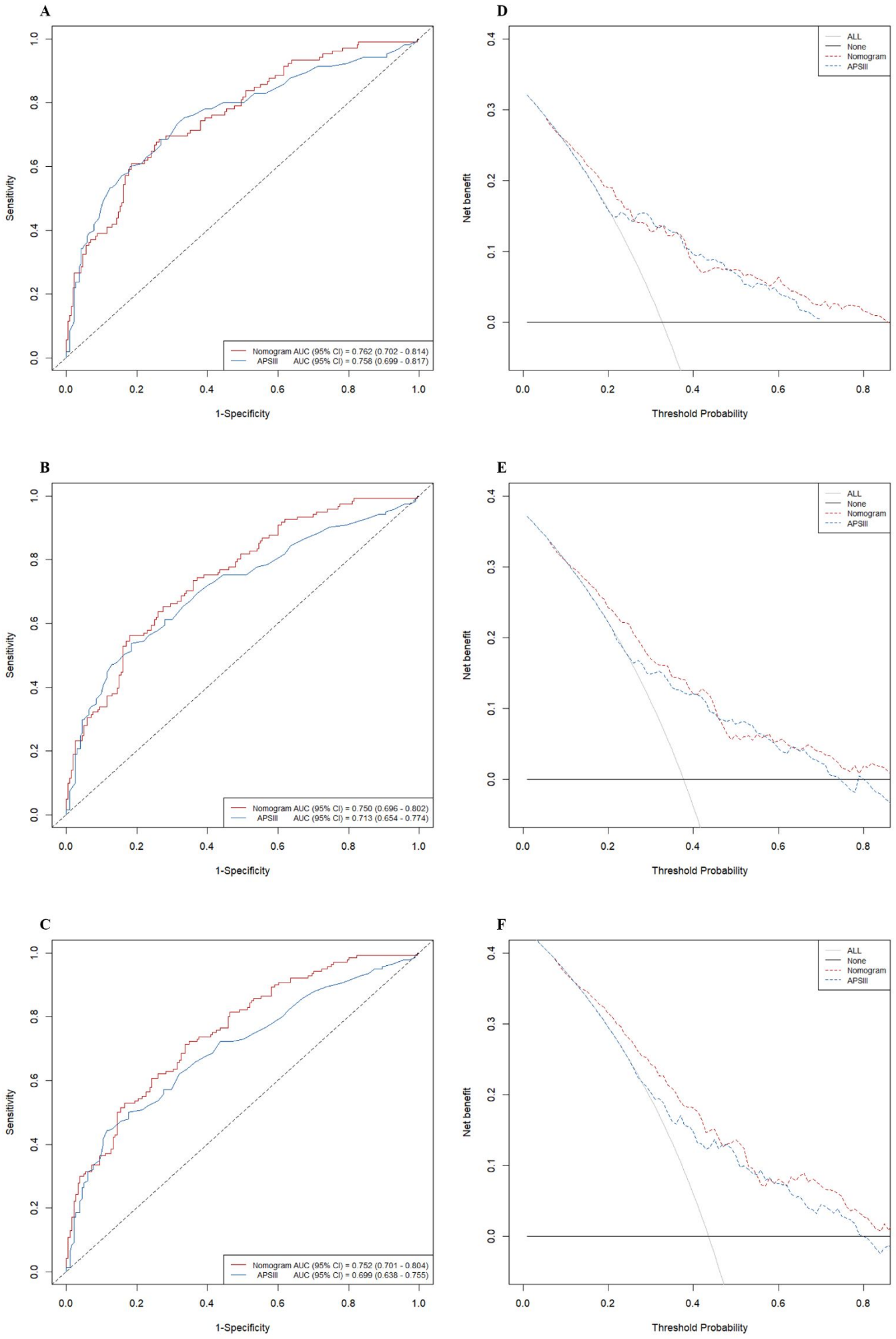


Figure 9. Time-dependent ROC curves using the nomogram and APSIII predict overall survival probabilities within 30-day **(A)**, 90-day **(B)**, and 180-day **(C)** in the

training set. Decision curve analysis of nomogram and APSIII for 30-day **(D)**, 90-day **(E)**, and 180-day **(F)** survival probabilities in the training set. APSIII: acute physiology score III; AUC: area under the curve; CI: confidence interval; ROC: receiver operating characteristic.

DISCUSSION

We have developed and validated a survival assessment model for pulmonary fibrosis that used commonly measured clinical and physiological variables to accurately predict the overall survival rate in patients with this condition. The model and grading system were based on easily calculable scores while maintaining the discriminative power of continuous models. Therefore, we proposed that this prediction model be used as a straightforward risk screening method for individuals diagnosed with pulmonary fibrosis. This will provide clinicians and patients with a structured framework to discuss prognosis, offer policymakers a tool to explore stage-specific management options, and enable researchers to identify high-risk study populations, such as determining whether early lung transplantation is advisable. Ultimately, our aim is to optimize the efficiency and effectiveness of clinical trials.

By analyzing data from 459 patients in the MIMIC-IV database and performing LASSO Cox regression and multivariate Cox regression analysis, we identified five independent risk factors: age, RDW, PO_2, temperature, and APSIII. Among these, APSIII had the strongest predictive power for mortality risk and is widely used to assess disease severity. However, our comprehensive model, when compared to the APSIII score alone, proved to be significantly better and statistically superior. In this study, we divided the total score into high risk and low risk groups using the median method, with 112.07 as the cut-off point. To enhance the robustness of our findings, Kaplan-Meier curves were used to reevaluate 30-day, 90-day, and 180-day mortality, showing significantly higher mortality rates in the high risk group compared to the low risk group at each time point.

While various prognostic models exist for pulmonary fibrosis, such as the Model for Gauging Mortality for Individual Patients, GAP Model, Longitudinal GAP Model, CT-GAP Model, and Modified GAP Model, most predict mortality within 1-3 years, primarily for idiopathic pulmonary fibrosis (IPF) patients [2, 11–14]. In contrast, our comprehensive model encompasses all types of pulmonary fibrosis in ICU admissions and predicts overall survival rate at 30-day, 90-day, and 180-day. Promptly assessing risk levels for ICU-admitted patients is crucial for timely treatment planning and effective communication with families.

In our model, age is a predictive factor for survival rate associated with pulmonary fibrosis, consistent with the previously mentioned five models. Numerous studies have demonstrated that aging impairs reparative mechanisms and hinders fibrosis regression in animal models of pulmonary fibrosis [15]. Advanced age is linked to a poorer prognosis in fibrotic interstitial lung diseases, such as chronic hypersensitivity pneumonitis (cHP), nonspecific interstitial pneumonia (NSIP), and organizing pneumonia (OP), compared to IPF [16]. It has been proposed that, in the context of IPF, epigenetic drift due to spontaneous stochastic errors in gene methylation accumulation with aging may trigger aberrant reactivation of embryonic developmental pathways, which interact with transforming growth factor-beta (TGF-β) to initiate and sustain the fibrotic feedback loop [17, 18]. Baumgartner KB et al. also showed that age is a risk factor for IPF [19]. In patients with IPF, numerous complications often arise, and elderly patients, who may already have a worse prognosis due to their age, are particularly susceptible to these complications [20–22]. The same holds true for individuals suffering from pulmonary fibrosis, this pattern is consistent with our model. Resting and exertional hypoxemia, assessed using the 6-minute walk test, has been associated with increased mortality in patients with IPF [23]. Our model also indicates that hypoxemia is a high-risk factor for death in patients with pulmonary fibrosis.

Unlike many existing models, we found that RDW is a risk factor for mortality in patients with pulmonary fibrosis. RDW is a parameter reflecting the degree of heterogeneity between red cells (anisocytosis) and is commonly used in hematology laboratories to classify anemia [24]. Recent evidence suggests that RDW is associated with complex disease prognosis, including malignant diseases [25, 26], thrombotic disease [27], liver disease, kidney failure [28, 29], community-acquired pneumonia, obstructive sleep apnea hypopnea syndrome [30], and cardiovascular disease [31]. It is also used in assessing the prognosis of patients with pulmonary hypertension [24]. Furthermore, RDW is associated with an increased risk of death in the general population [32]. Another study confirmed that RDW, along with conditions like IPF and COPD, is associated with poor clinical outcomes [33–35]. Although the specific mechanism remains unclear, some studies suggest that increased RDW may serve as a biomarker for early hypoxemia [33, 36, 37]. Arterial hypoxemia stimulates increased erythropoietin secretion, affecting erythrocyte maturation and survival, thus elevating RDW [33, 36]. Elevated RDW might help in the early identification of IPF patients with intermittent hypoxemia [38]. One study reported that patients with RDW >15% had lower median diffusion capacity of lung for carbon monoxide %predicted (DLCO%pred) and an increased risk of death compared to those with RDW ≤15% [38]. Similar results for DLCO%pred were found in the study by Karampitsakos et al. [39]. Some researchers have demonstrated a strong correlation between RDW and PaO_2, indicating that high RDW levels may signal hypoxia. Decreased PaO_2 levels trigger erythropoietin release, resulting in alterations in red cell size distribution. RDW levels tend to rise with the severity of lung disease, suggesting its potential as a predictor of disease severity and mortality [40]. However, the pathophysiological basis of the association between RDW and IPF mortality remains unclear. Elevated RDW values may stem from an underlying inflammatory state that affects erythropoiesis, RBC circulating

half-life, and RBC membrane deformability. The possibility of underlying pulmonary hypertension contributing to increased RDW in the high RDW group may explain the higher mortality in this group [38].

Our model indicates that body temperature is a significant factor for survival, with the nomogram illustrating that as body temperature decreases, patient prognosis worsens. However, mild hypothermia has been shown to reduce lung injury and enhance lung function during acute respiratory distress syndrome [41]. Wei F et al. suggested that lowering whole-body temperature to 33-35°C using cryogenic fluorocarbon partial fluid ventilation could improve respiratory function, decrease acute lung injury induced inflammatory factors, and enhance anti-inflammatory responses [42]. These findings seem to conflict with our study results. Several factors might explain this discrepancy: Firstly, hypothermia has long been suspected to increase the risk of infectious pulmonary complications [43, 44]. In a clinical study, Perbet et al. demonstrated that therapeutic hypothermia was an independent risk factor for early-onset pneumonia [45]. Secondly, when body temperature drops, vasoconstriction and bradycardia occur, affecting perfusion and oxygen delivery to vital organs [46]. Additionally, hypothermia can lead to abnormal coagulation, exacerbating the condition [47].

This study also has several limitations: First, our data comes from a single academic center and is based on a retrospective study design using a public database, which presents certain limitations. The database does not provide specific reasons for patients' hospital admissions or ICU transfers, only the final diagnosis. We attempted various data extraction methods at the beginning of the study to address this issue, but were unsuccessful. This is the main limitation of our study. Second, the MIMIC-IV database provides only the time of death for deceased patients and lacks detailed information on the cause of death. This limitation hinders our ability to determine whether the deaths were related to PF or other competing risks. This is another significant limitation of our study. Additionally, we did not include indicators reflecting the severity of the respiratory system in patients with PF, such as FiO_2, respiratory support methods, pulmonary function tests, and CT scans, primarily because the proportion of missing data for these indicators was too high or they were unavailable. We emphasized that the model's utility is more relevant for ICU patients with PF generally, rather than those with acute respiratory issues. Validating our risk classification system through a prospective, multicenter study with larger sample sizes of critically ill PF patients is imperative before clinical implementation. Developing models based on extensive datasets from multiple centers is a viable approach to address the challenge of limited data. We anticipate that these measures will enhance the robustness and applicability of our model.

Conclusion

This study has successfully developed and validated a prediction model that integrates five common clinical features, enabling effective prediction of 30-day, 90-day, and 180-day OS in ICU-admitted patients with PF for any reasons. The included indicators are straightforward to obtain, making the model valuable for clinicians in formulating tailored treatment plans and communication strategies with the families of pulmonary fibrosis patients early on.

Abbreviations

LASSO: least absolute selection and shrinkage operator; ROC: receiver operating characteristic; DCA : decision curve analysis; MIMIC-IV: Medical Information Mart for Intensive Care IV; ICU: Intensive Care Unit; MAP: mean arterial blood pressure; RR: respiratory rate; RBC: red blood cell; WBC: white blood cell; RDW: red cell distribution width; BUN: blood urea nitrogen; PaO_2: partial pressure of arterial oxygen; $PaCO_2$: carbon dioxide partial pressure; APTT: activated partial thromboplastin time; PT: prothrombin time; INR: international normalized ratio; SIRS: systemic inflammatory response syndrome; APSIII: acute physiology score III; SOFA: sequential organ failure assessment; GCS: Glasgow coma scale; SAPSII: simplified acute physiology score II; OASIS: Oxford acute severity of illness score; LODS : logistic organ dysfunction system; COPD: chronic obstructive pulmonary disease; OSAS: obstructive sleep apnea syndrome; SD: standard deviation; IQR: interquartile range; OS: overall survival; VIF: variance inflation factor; RCS: restricted cubic spline; HR: hazard ratio; CI: confidence interval; IPF: idiopathic pulmonary fibrosis.

Acknowledgements

We acknowledged the contributions of the MIMIC-IV (version 2.2) program registry for creating and updating the MIMIC IV database.

Author contributions

Ruxian Sun designed the study. Ruxian Sun extracted and analyzed the data. Yuanjun Zhou and Changyan Tao drafted the manuscript. Juan Xu and Jiao Lv revised the manuscript. The final version was approved by all authors.

Funding

None.

Availability of data and materials

The datasets analysed during the current study are available in the MIMIC-IV repository, https://physionet.org/content/mimiciv/2.2/.

Declarations

Ethics approval and consent to participate

The MIMIC-IV database received approval from the Institutional Review Boards of both Beth Israel Deaconess Medical Center and the Massachusetts Institute of Technology, with informed consent obtained from each patient. In this study, local institutional review board approval and informed consent were not required because the MIMIC research data were publicly available and all patient data were de-identified.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Reference

1. Maher TM. Interstitial lung disease. Interstitial Lung Dis. 2024.
2. Ley B, Ryerson CJ, Vittinghoff E, Ryu JH, Tomassetti S, Lee JS, et al. A multidimensional index and staging system for idiopathic pulmonary fibrosis. Ann Intern Med. 2012;156:684–91.
3. Blazek K, Van Zwieten A, Saglimbene V, Teixeira-Pinto A. A practical guide to multiple imputation of missing data in nephrology. Kidney Int. 2021;99:68–74.
4. Austin PC, White IR, Lee DS, Van Buuren S. Missing data in clinical research: a tutorial on multiple imputation. Can J Cardiol. 2021;37:1322–31.
5. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010;33.
6. Hu R, Zhou S, Liu Y, Tang Z. Margin-based pareto ensemble pruning: An ensemble pruning algorithm that learns to search optimized ensembles. Comput Intell Neurosci. 2019;2019:1–12.
7. Tang Z, Chen Y, Wang Z, Hu R, Wu EQ. Non-spike timing-dependent plasticity learning mechanism for memristive neural networks. Appl Intell. 2021;51:3684–95.
8. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: More than meets the eye. Lancet Oncol. 2015;16:e173-180.
9. Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. BMC Med Inform Decis Mak. 2008;8:53.
10. Localio AR, Goodman S. Beyond the usual prediction accuracy metrics: Reporting results for clinical decision making. Ann Intern Med. 2012;157:294.
11. du Bois RM, Weycker D, Albera C, Bradford WZ, Costabel U, Kartashov A, et al. Ascertainment of individual risk of mortality for patients with idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2011;184:459–66.
12. Ley B, Bradford WZ, Weycker D, Vittinghoff E, du Bois RM, Collard HR. Unified baseline and longitudinal mortality prediction in idiopathic pulmonary fibrosis. Eur Respir J. 2015;45:1374–81.
13. Ley B, Elicker BM, Hartman TE, Ryerson CJ, Vittinghoff E, Ryu JH, et al. Idiopathic pulmonary fibrosis: CT and risk of death. Radiology. 2014;273:570–9.
14. Ryerson CJ, Vittinghoff E, Ley B, Lee JS, Mooney JJ, Jones KD, et al. Predicting survival across chronic interstitial lung disease: The ILD-GAP model. Chest. 2014;145:723–8.
15. Hecker L, Logsdon NJ, Kurundkar D, Kurundkar A, Bernard K, Hock T, et al. Reversal of persistent fibrosis in aging by targeting Nox4-Nrf2 redox imbalance. Sci Transl Med. 2014;6:231ra47.
16. Selman M, Pardo A. Revealing the pathogenic and aging-related mechanisms of the enigmatic idiopathic pulmonary fibrosis. an integral model. Am J Respir Crit Care Med. 2014;189:1161–72.
17. Selman M, López-Otín C, Pardo A. Age-driven developmental drift in the pathogenesis of idiopathic pulmonary fibrosis. Eur Respir J. 2016;48:538–52.
18. Placido L, Romero Y, Maldonado M, Toscano-Marquez F, Ramírez R, Calyeca J, et al. Loss of MT1-MMP in alveolar epithelial cells exacerbates pulmonary fibrosis. Int J Mol Sci. 2021;22:2923.
19. Baumgartner KB, Samet JM, Stidley CA, Colby TV, Waldron JA. Cigarette smoking: A risk factor for idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 1997;155:242–8.
20. Raghu G, Amatto VC, Behr J, Stowasser S. Comorbidities in idiopathic pulmonary fibrosis patients: A systematic literature review. Eur Respir J. 2015;46:1113–30.
21. Meyer KC, Danoff SK, Lancaster LH, Nathan SD. Management of idiopathic pulmonary fibrosis in the elderly patient: Addressing key questions. Chest. 2015;148:242–52.
22. King CS, Nathan SD. Idiopathic pulmonary fibrosis: Effects and optimal management of comorbidities. Lancet Respir Med. 2017;5:72–84.
23. Khor YH, Gutman L, Abu Hussein N, Johannson KA, Glaspole IN, Guler SA, et al. Incidence and prognostic significance of hypoxemia in fibrotic interstitial lung disease: An international cohort study. Chest. 2021;160:994–1005.
24. Baltazares-Lipp ME, Aguilera-Velasco A, Aquino-Gálvez A, Velázquez-Cruz R, Hernández-Zenteno RJ, Alvarado-Vásquez N, et al. Evaluating of red blood cell distribution width, comorbidities and electrocardiographic ratios as predictors of prognosis in patients with pulmonary hypertension. Diagn Basel Switz. 2021;11:1297.
25. Zorlu A, Bektasoglu G, Guven FMK, Dogan OT, Gucuk E, Ege MR, et al. Usefulness of admission red cell distribution width as a predictor of early mortality in patients with acute pulmonary embolism. Am J Cardiol. 2012;109:128–34.
26. Koma Y, Onishi A, Matsuoka H, Oda N, Yokota N, Matsumoto Y, et al. Increased red blood cell distribution width associates with cancer stage and prognosis in patients with lung cancer. PloS One. 2013;8:e80240.
27. Rezende SM, Lijfering WM, Rosendaal FR, Cannegieter SC. Hematologic variables and venous thrombosis: Red cell distribution width and blood monocyte count are associated with an increased risk. Haematologica. 2014;99:194–200.
28. Turcato G, Campagnaro T, Bonora A, Vignola N, Salvagno GL, Cervellin G, et al. Red blood cell distribution width independently predicts 1-month mortality in acute decompensation of cirrhotic patients admitted to emergency department. Eur J Gastroenterol Hepatol. 2018;30:33–8.
29. Wang B, Lu H, Gong Y, Ying B, Cheng B. The association between red blood cell distribution width and mortality in critically ill patients with acute kidney injury. BioMed Res Int. 2018;2018:9658216.
30. Bello S, Fandos S, Lasierra AB, Mincholé E, Panadero C, Simon AL, et al. Red blood cell distribution width [RDW] and long-term mortality after community-acquired pneumonia. A comparison with proadrenomedullin. Respir Med. 2015;109:1193–206.
31. Danese E, Lippi G, Montagnana M. Red blood cell distribution width and cardiovascular diseases. J Thorac Dis. 2015;7:E402-411.
32. Perlstein TS, Weuve J, Pfeffer MA, Beckman JA. Red blood cell distribution width and mortality risk in a community-based prospective cohort. Arch Intern Med. 2009;169:588–94.
33. Karampitsakos T, Dimakou K, Papaioannou O, Chrysikos S, Kaponi M, Bouros D, et al. The role of increased red cell distribution width as a negative prognostic marker in patients with COPD. Pulm Pharmacol Ther. 2020;60:101877.
34. Hampole CV, Mehrotra AK, Thenappan T, Gomberg-Maitland M, Shah SJ. Usefulness of red cell distribution width as a prognostic marker in pulmonary hypertension. Am J Cardiol. 2009;104:868–72.
35. Braun E, Domany E, Kenig Y, Mazor Y, Makhoul BF, Azzam ZS. Elevated red cell distribution width predicts poor outcome in young patients with community acquired pneumonia. Crit Care Lond Engl. 2011;15:R194.
36. Epstein D, Nasser R, Mashiach T, Azzam ZS, Berger G. Increased red cell distribution width: A novel predictor of adverse outcome in patients hospitalized due to acute exacerbation of chronic obstructive pulmonary disease. Respir Med. 2018;136:1–7.
37. Karampitsakos T, Akinosoglou K, Papaioannou O, Panou V, Koromilias A, Bakakos P, et al. Increased red cell distribution width is associated with disease severity in hospitalized adults with SARS-CoV-2 infection: An observational multicentric study. Front Med. 2020;7:616292.
38. Nathan SD, Reffett T, Brown AW, Fischer CP, Shlobin OA, Ahmad S, et al. The red cell distribution width as a prognostic indicator in idiopathic pulmonary fibrosis. Chest. 2013;143:1692–8.
39. Karampitsakos T, Torrisi S, Antoniou K, Manali E, Korbila I, Papaioannou O, et al. Increased monocyte count and red cell distribution width as prognostic biomarkers in patients with idiopathic pulmonary fibrosis. Respir Res. 2021;22:140.
40. Gürün Kaya A, Özyürek BA, Şahin Özdemirel T, Öz M, Erdoğan Y. Prognostic significance of red cell distribution width in idiopathic pulmonary fibrosis and combined pulmonary fibrosis emphysema. Med Princ Pract Int J Kuwait Univ Health Sci Cent. 2021;30:154–9.
41. Correction to lancet respir med 2020; published online feb 21. https://doi.org/10.1016/S2213-2600(20)30079-5. Lancet Respir Med. 2020;8:e26.
42. Wei F, Hu Y, Jiang M, Ye L, Yang L. Effect of perfluorocarbon partial liquid ventilation-induced hypothermia on dogs with acute lung injury. Ann Palliat Med. 2020;9:2141–51.
43. Mongardon N, Perbet S, Lemiale V, Dumas F, Poupet H, Charpentier J, et al. Infectious complications in out-of-hospital cardiac arrest patients in the therapeutic hypothermia era. Crit Care Med. 2011;39:1359–64.
44. Yanagawa Y, Ishihara S, Norio H, Takino M, Kawakami M, Takasu A, et al. Preliminary clinical outcome study of mild resuscitative hypothermia after out-of-

hospital cardiopulmonary arrest. Resuscitation. 1998;39:61–6.
45. Perbet S, Mongardon N, Dumas F, Bruel C, Lemiale V, Mourvillier B, et al. Early-onset pneumonia after cardiac arrest: Characteristics, risk factors and influence on prognosis. Am J Respir Crit Care Med. 2011;184:1048–54.
46. Boorman LW, Harris SS, Shabir O, Lee L, Eyre B, Howarth C, et al. Bidirectional alterations in brain temperature profoundly modulate spatiotemporal neurovascular responses in-vivo. Commun Biol. 2023;6:185.
47. Chen X, Shang W, Huang X, Shu L, Xiao S, Jiang Q, et al. The effect of winter temperature on patients with ischemic stroke. Med Sci Monit Int Med J Exp Clin Res. 2019;25:3839–45.

Table 1. Demographic and clinical characteristics of the training and validation sets

Variables	Total (n = 459)	Training set (n = 321)	Validation set (n = 138)	P value
Demographics				
Age, years	76 (66-83)	76 (67-82)	76 (66-84)	0.754
Female, n (%)	214 (46.62%)	142 (44.24%)	72 (52.17%)	0.118
Weight, kg	34.4 (28.9-40.0)	34.2 (28.6-39.7)	34.6 (29.3-40.4)	0.554
Smoker, n (%)	105 (22.88%)	74 (23.05%)	31 (22.46%)	0.890
Vital signs				
HR, beats/min	89 (76-103)	88 (76-102)	89 (76-104)	0.321
MAP, mmHg	80 (71-92)	80 (71-92)	79 (71-91)	0.909
RR, times/min	20 (16-26)	20 (16-26)	20 (17-26)	0.593
Temperature, ℃	36.7 (36.4-36.9)	36.7 (36.4-36.9)	36.7 (36.3-36.9)	0.731
Laboratory tests				
RBC, m/uL	3.71±0.76	3.72±0.73	3.70±0.81	0.806
WBC, K/uL	11.7 (8.5-15.7)	11.2 (8.3-15.2)	11.9 (8.8-16.0)	0.333
Hemoglobin, g/dL	11.12±2.22	11.13±2.14	11.09±2.43	0.852
Hematocrit, %	33.99±6.59	34.09±6.34	33.74±7.14	0.600
RDW, %	14.8 (13.7-16.1)	14.7 (13.7-16.1)	14.8 (13.7-16.1)	0.844
Platelet, K/uL	227 (161-307)	227 (159-302)	227 (172-316)	0.446
Creatinine, mg/dL	1.0 (0.8-1.4)	1.0 (0.8-1.4)	1.0 (0.8-1.3)	0.856
BUN, mg/dL	21 (15-32)	21 (15-32)	22 (15-31)	0.471
Potassium, mEq/L	4.2 (3.8-4.6)	4.2 (3.8-4.6)	4.2 (3.9-4.6)	0.553
Sodium, mEq/L	138 (135-140)	138 (135-141)	138 (135-140)	0.368
Chloride, mEq/L	101 (97-106)	101 (97-106)	101 (98-105)	0.837
Calcium, mg/dL	8.5 (8.1-9.0)	8.5 (8.1-9.0)	8.5 (8.0-8.9)	0.665
Blood glucose, mg/dL	128 (105-169)	127 (104-166)	134 (107-173)	0.109
PH	7.40 (7.34-7.45)	7.40 (7.33-7.46)	7.40 (7.35-7.44)	0.915
PaO_2, mmHg	86 (52-161)	88 (53-160)	80 (51-172)	0.505
$PaCO_2$, mmHg	42 (36-49)	43 (36-49)	41 (37-46)	0.573
Lactic acid, mmol/L	1.7 (1.2-2.5)	1.7 (1.2-2.5)	1.7 (1.2-2.5)	0.885
Hydrocarbonate, mEq/L	24 (22-27)	24 (22-27)	24 (22-27)	0.677
APTT, s	30.0 (26.8-35.9)	30.3 (26.9-35.9)	29.6 (26.4-35.7)	0.569
PT, s	13.6 (12.4-16.2)	13.6 (12.5-16.3)	13.6 (12.2-15.9)	0.588
INR	1.2 (1.1-1.5)	1.2 (1.1-1.5)	1.2 (1.1-1.4)	0.477
24-hour urine volume, ml	1400 (900-2265)	1395 (875-2170)	1449 (1001-2571)	0.157
Clinical score				
SIRS	3 (2-3)	3 (2-3)	3 (2-3)	0.940
APSIII	44 (34-58)	43 (34-58)	45 (34-57)	0.790
SOFA	4 (2-6)	4 (2-6)	4 (2-6)	0.201
GCS	14 (13-15)	14 (13-15)	14 (13-15)	0.473
SAPSII	35 (28-42)	35 (28-43)	35 (28-41)	0.990
OASIS	32 (27-38)	32 (27-38)	32 (26-37)	0.246
LODS	4 (2-6)	4 (2-7)	4 (2-6)	0.589
Comorbidities, n (%)				
Myocardial infarct	78 (16.99%)	56 (17.45%)	22 (15.94%)	0.694

Congestive heart failure	209 (45.53%)	146 (45.48%)	63 (45.65%)	0.973
Peripheral vascular disease	65 (14.16%)	47 (14.64%)	18 (13.04%)	0.652
Cerebrovascular disease	48 (10.46%)	37 (11.53%)	11 (7.97%)	0.254
COPD	237 (51.63%)	170 (52.96%)	67 (48.55%)	0.386
Rheumatic disease	47 (10.24%)	33 (10.28%)	14 (10.14%)	0.965
Renal disease	109 (23.75%)	79 (24.61%)	30 (21.74%)	0.507
Malignant cancer	56 (12.20%)	39 (12.15%)	17 (12.32%)	0.959
Atrial fibrillation	153 (33.33%)	112 (34.89%)	41 (29.71%)	0.280
Hypertension	311 (67.76%)	217 (67.60%)	94 (68.12%)	0.914
Asthma	48 (10.46%)	36 (11.21%)	12 (8.70%)	0.419
Pulmonary arterial hypertension	25 (5.45%)	17 (5.30%)	8 (5.80%)	0.828
OSAS	55 (11.98%)	38 (11.84%)	17 (12.32%)	0.884
Esophageal reflux	129 (28.10%)	89 (27.73%)	40 (28.99%)	0.783
Osteoporosis	39 (8.50%)	30 (9.35%)	9 (6.52%)	0.320
Diabetes	136 (29.63%)	101 (31.46%)	35 (25.36%)	0.189
Liver disease	40 (8.71%)	30 (9.35%)	10 (7.25%)	0.465
Charlson comorbidity index	6 (5-8)	6 (5-8)	6 (5-8)	0.209
Outcome				
30-day OS rate, n (%)	314 (68.41%)	98 (71.01%)	216 (67.29%)	0.431
90-day OS rate, n (%)	288 (62.75%)	88 (63.77%)	200 (62.31%)	0.766
180-day OS rate, n (%)	259 (56.43%)	78 (56.52%)	181 (56.39%)	0.979

Continuous variables are presented as mean ± SD if normally distributed, and median (interquartile range) if not normally distributed. Categorical variables are presented as number of patients (%). HR: heart rate; MAP: mean arterial blood pressure; RR: respiratory rate; RBC: red blood cell; WBC: white blood cell; RDW: red cell distribution width; BUN: blood urea nitrogen; PH: potential of hydrogen; PaO_2: partial pressure of arterial oxygen; $PaCO_2$: carbon dioxide partial pressure; APTT: activated partial thromboplastin time; PT: prothrombin time; INR: international normalized ratio; SIRS: systemic inflammatory response syndrome; APSIII: acute physiology score III; SOFA: sequential organ failure assessment; GCS: Glasgow coma scale; SAPSII: simplified acute physiology score II; OASIS: oxford acute severity of illness score; LODS: logistic organ dysfunction system; COPD: chronic obstructive pulmonary disease; OSAS: obstructive sleep apnea syndrome; OS: overall survival.

Table 2. Multivariate cox regression analysis based on LASSO regression results

Variables	Coefficients	HR (95%CI)	P value
Age	0.012156	1.0122 (1.0015-1.0246)	**0.0493**
Temperature	-0.229618	0.7948 (0.6582-0.9599)	**0.0171**
WBC	0.023057	1.0233 (0.9990-1.0483)	0.0605
RDW	0.079686	1.0829 (1.0101-1.1610)	**0.0249**
PaO_2	-0.003471	0.9965 (0.9950-0.9981)	**<0.0001**
SIRS	0.151512	1.1636 (0.9786-1.3836)	0.0864
APSIII	0.010508	1.0106 (1.0003-1.0209)	**0.0434**
OASIS	0.013335	1.0134 (0.9907-1.0367)	0.2495
LODS	0.055595	1.0572 (0.9789-1.1417)	0.1566

LASSO: least absolute shrinkage and selection operator; HR: hazard ratio; CI: confidence interval; WBC: white blood cell;

RDW: red cell distribution width; PaO_2: partial pressure of arterial oxygen; SIRS: systemic inflammatory response syndrome; APSIII: acute physiology score III; OASIS: oxford acute severity of illness score; LODS: logistic organ dysfunction system.

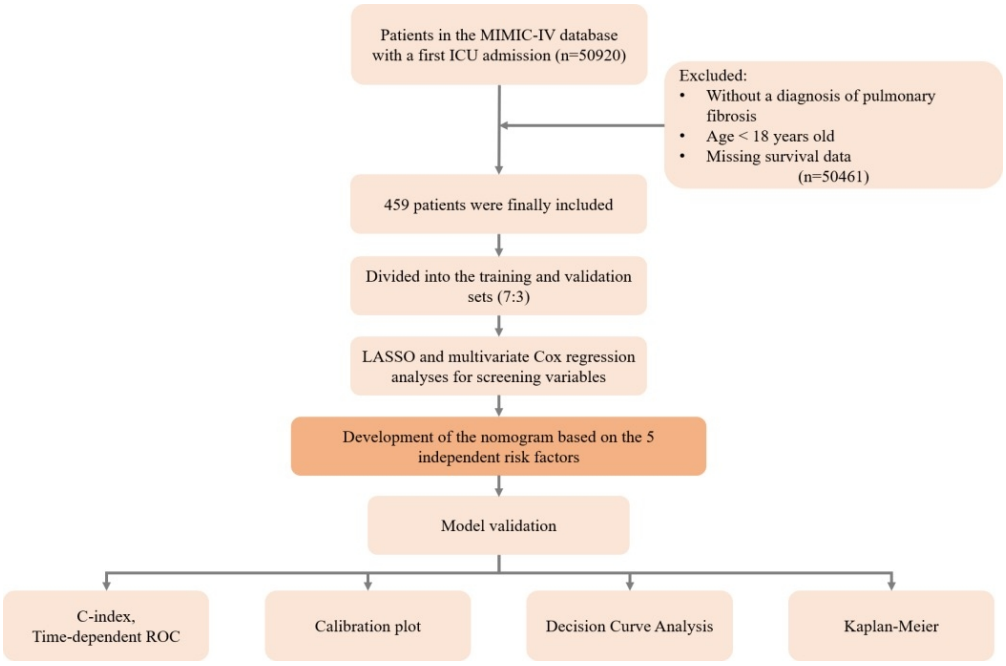


Figure 1. Workflow of the study. MIMIC-IV: Medical Information Mart for Intensive Care IV; ICU: intensive care unit; LASSO: least absolute shrinkage and selection operator. ROC: receiver operating characteristic.

177x116mm (150 x 150 DPI)

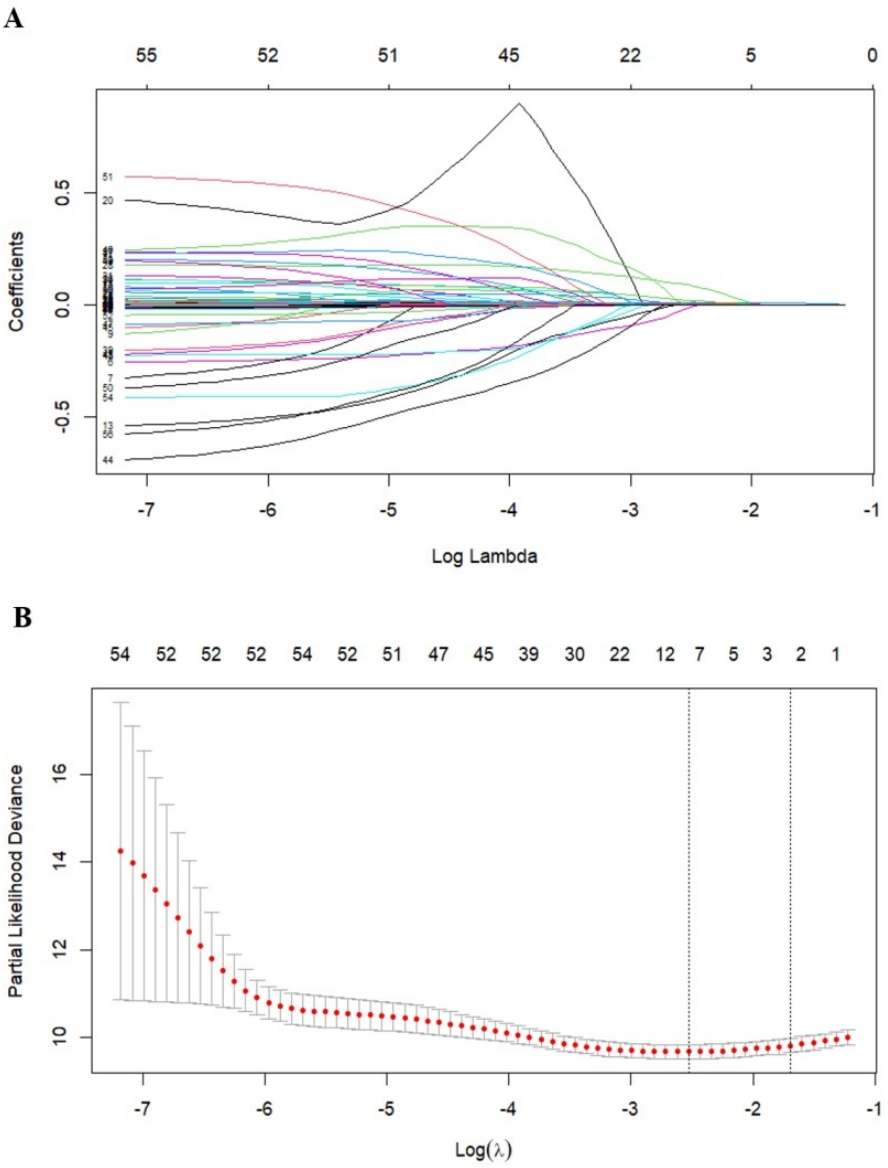


Figure 2. The LASSO regression analysis to select the potential variables. A total of fifty-five variables were initially included and nine variables were finally selected for further analysis. The LASSO coefficient analysis of the clinical features (A). Cross-validation was conducted using a 10-fold approach to optimize the tuning parameter selection in the LASSO model (B). LASSO: least absolute shrinkage and selection operator.

144x184mm (150 x 150 DPI)

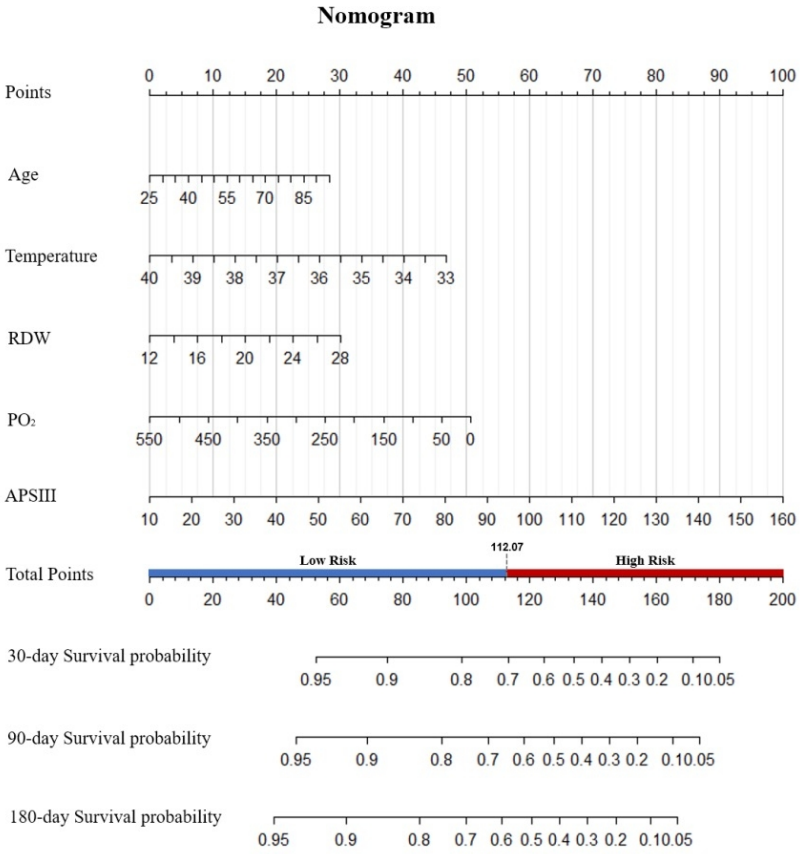


Figure 3. The nomogram for predicting 30-day, 90-day, and 180-day survival probabilities in patients with pulmonary fibrosis.

177x177mm (150 x 150 DPI)

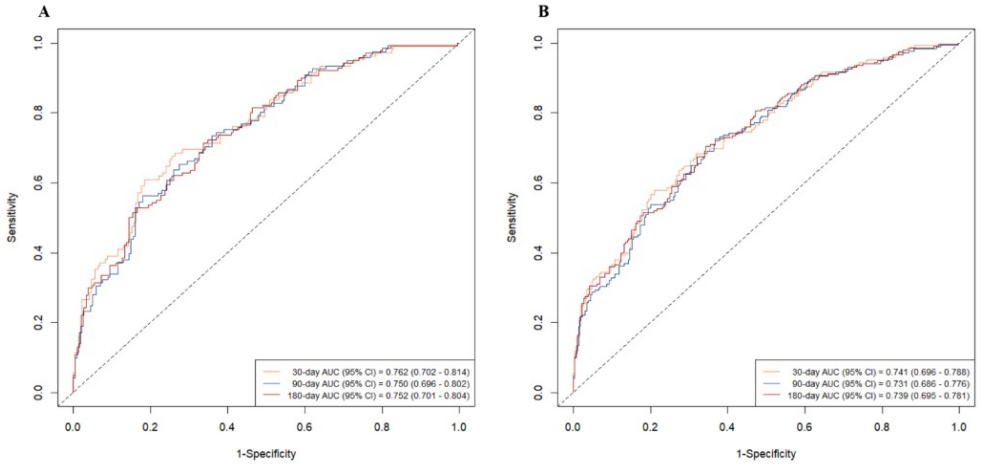


Figure 4. Time-dependent ROC curves using the nomogram predict overall survival probabilities within 30-day, 90-day, and 180-day in the training set (A) and validation set (B). AUC: area under the curve; CI: confidence interval; ROC: receiver operating characteristic.

120x60mm (220 x 220 DPI)

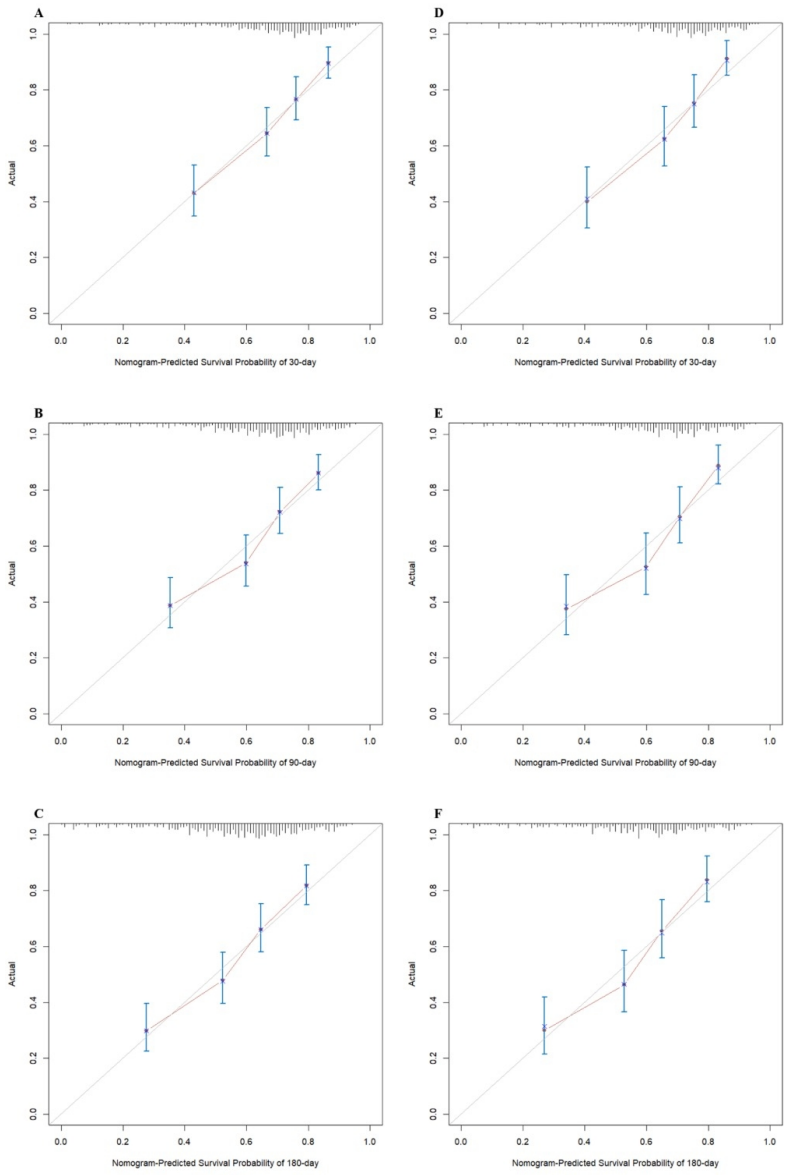
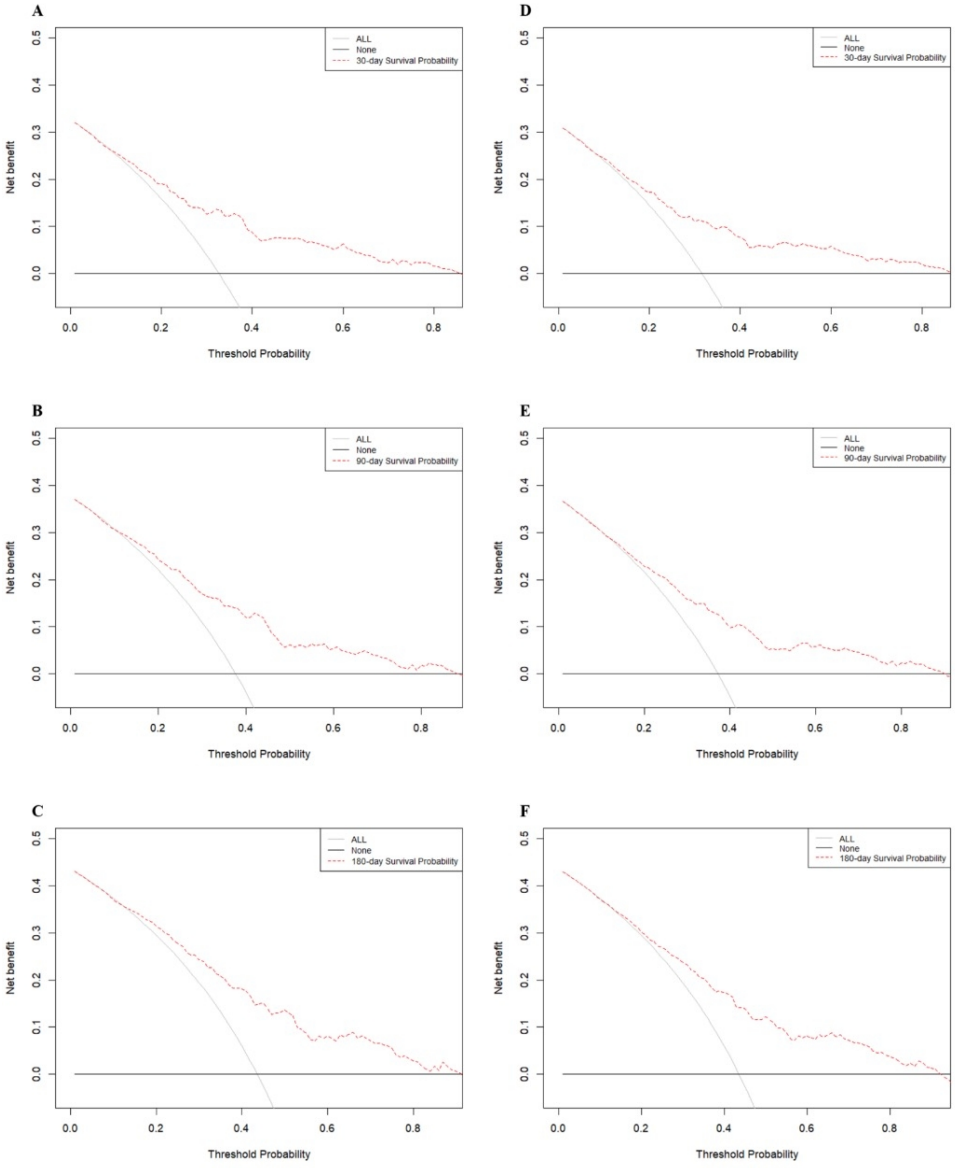


Figure 5. Calibration curves for OS at 30-day, 90-day, and 180-day in patients with pulmonary fibrosis. 30-day OS rate in the training set (A); 90-day OS rate in the training set (B); 180-day OS rate in the training set (C); 30-day OS rate in the validation set (D); 90-day OS rate in the validation set (E); 180-day OS rate in the validation set (F). The horizontal axis depicts the survival rate predicted by the nomogram, while the vertical axis represents the actual survival rate. The grey diagonal line indicates that the predicted survival rate perfectly matches the actual survival rate. OS: overall survival.

118x177mm (220 x 220 DPI)



Decision curve analysis of the training set for 30-day (A), 90-day (B), and 180-day (C) survival probabilities. Decision curve analysis of the validation set for 30-day (D), 90-day (E), and 180-day (F) survival probabilities. The y-axis measures the net benefit. The gray line represents the assumption that all patients are deceased. The black line represents the assumption of no patient deaths. The red dotted line represents the nomogram.

120x148mm (220 x 220 DPI)

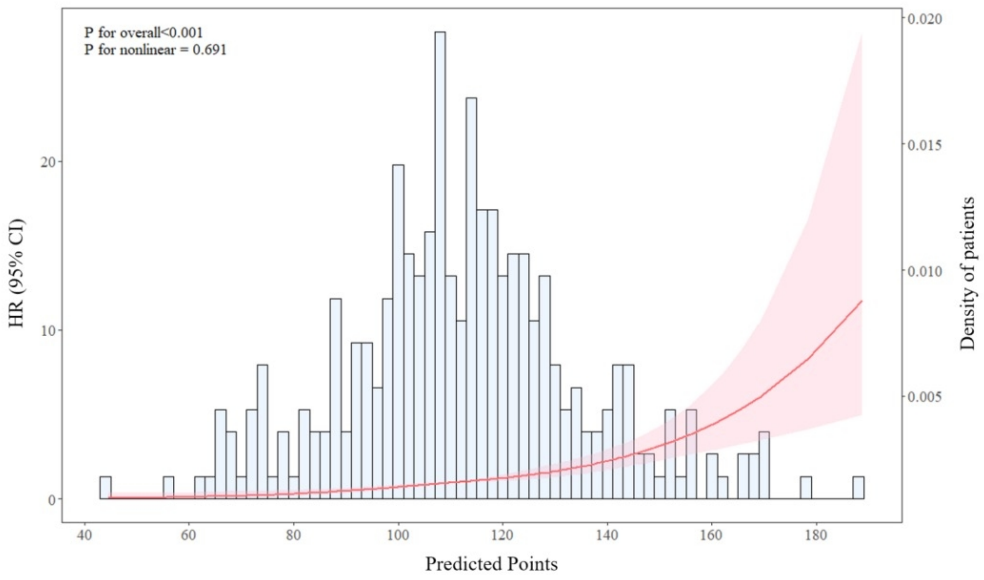


Figure 7. Restricted cubic spline regression analysis for the predicted points derived from the nomogram. The heavy central lines represent the estimated hazard ratios, with shaded ribbons indicating the corresponding 95% confidence intervals. The histogram illustrates the density of patients. HR: hazard ratio; CI: confidence interval.

120x73mm (220 x 220 DPI)

Dear Editor:

On behalf of our research team, I am submitting our latest research findings to your journal, Science Progress, The title is《Development and Validation of a Nomogram to Predict Survival in Pulmonary Fibrosis Patients Admitted to the ICU》。Pulmonary interstitial fibrosis, a common end-stage manifestation of interstitial pneumonia, is a chronic and progressive lung disease marked by inflammatory cell infiltration and extensive fibrosis in the lung interstitium. Prompt identification of high-risk patients is crucial for guiding clinical decisions and determining the optimal timing for lung transplantation. This study aimed to develop and validate a nomogram to predict the overall survival rate of patients with pulmonary fibrosis (PF) in the intensive care unit (ICU).

We believe that this research holds significant theoretical value and practical significance for the clinical diagnosis, treatment, and public health prevention and control of tuberculosis and respiratory diseases.

Here, we hope to present our research findings to experts and scholars in related fields both domestically and internationally through your esteemed journal, and share our research experiences and ideas with our peers. We are confident that this article will contribute to the development of the field of respiratory medicine and provide new perspectives to address clinical challenges in related areas.

This study has not been published in any other journal, nor has it been simultaneously submitted to other journals. We have adhered to all ethical review standards and ensured the authenticity and reproducibility of the data.

Thank you for taking the time to review our manuscript. We look forward to receiving your valuable feedback and hope for the opportunity to publish in your journal. Should you have any questions or require further information, please do not hesitate to let us know.

Thank you again for your consideration and support!

Sincerely,

YUANJUN ZHOU

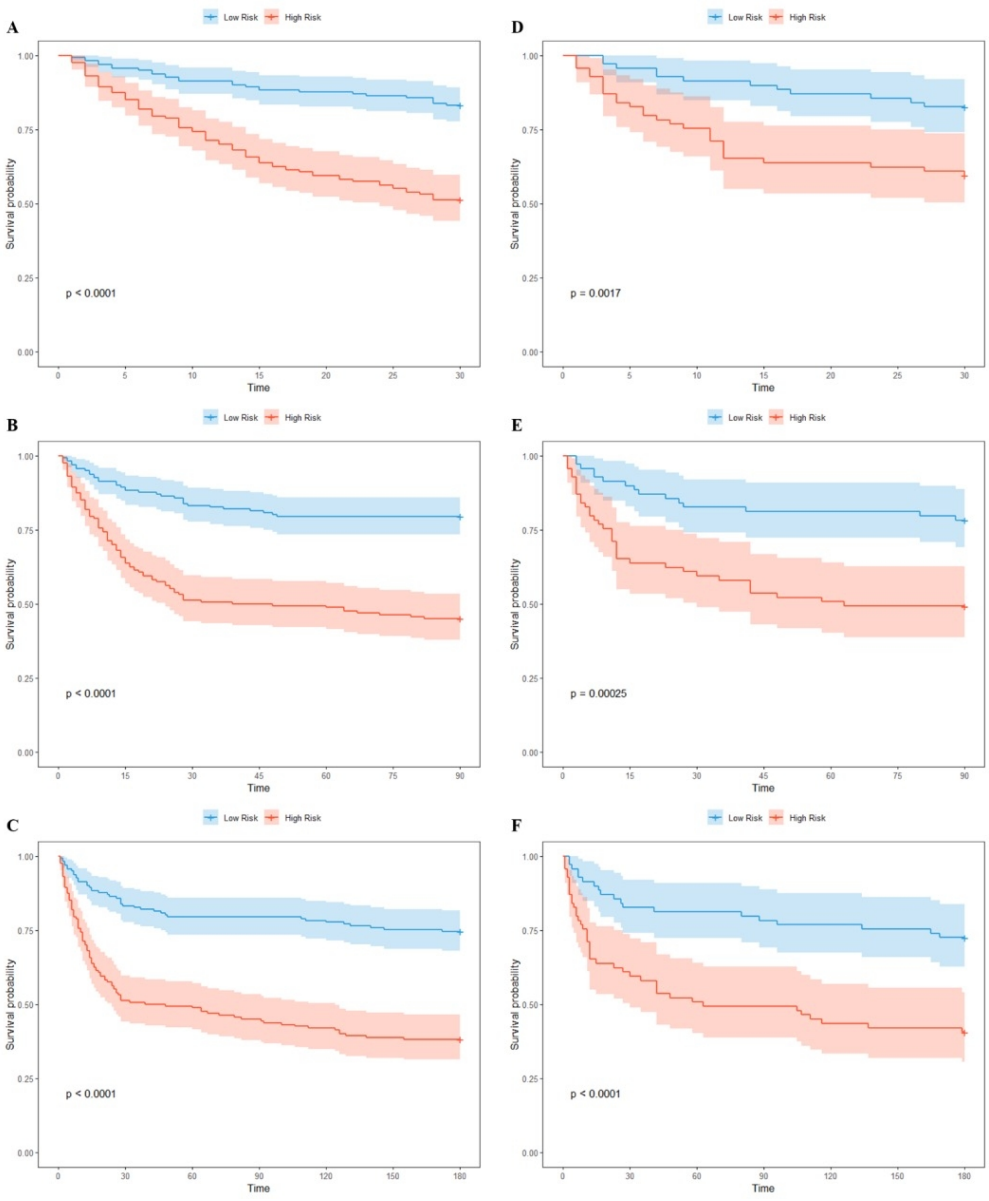


Figure 8. Kaplan-Meier survival curves of OS for patients with pulmonary fibrosis at 30-day (A), 90-day (B), and 180-day (C) in the training set, and at 30-day (D), 90-day (E), and 180-day (F) in the validation set. Patients with low risk were associated with better survival. OS: overall survival.

120x145mm (220 x 220 DPI)

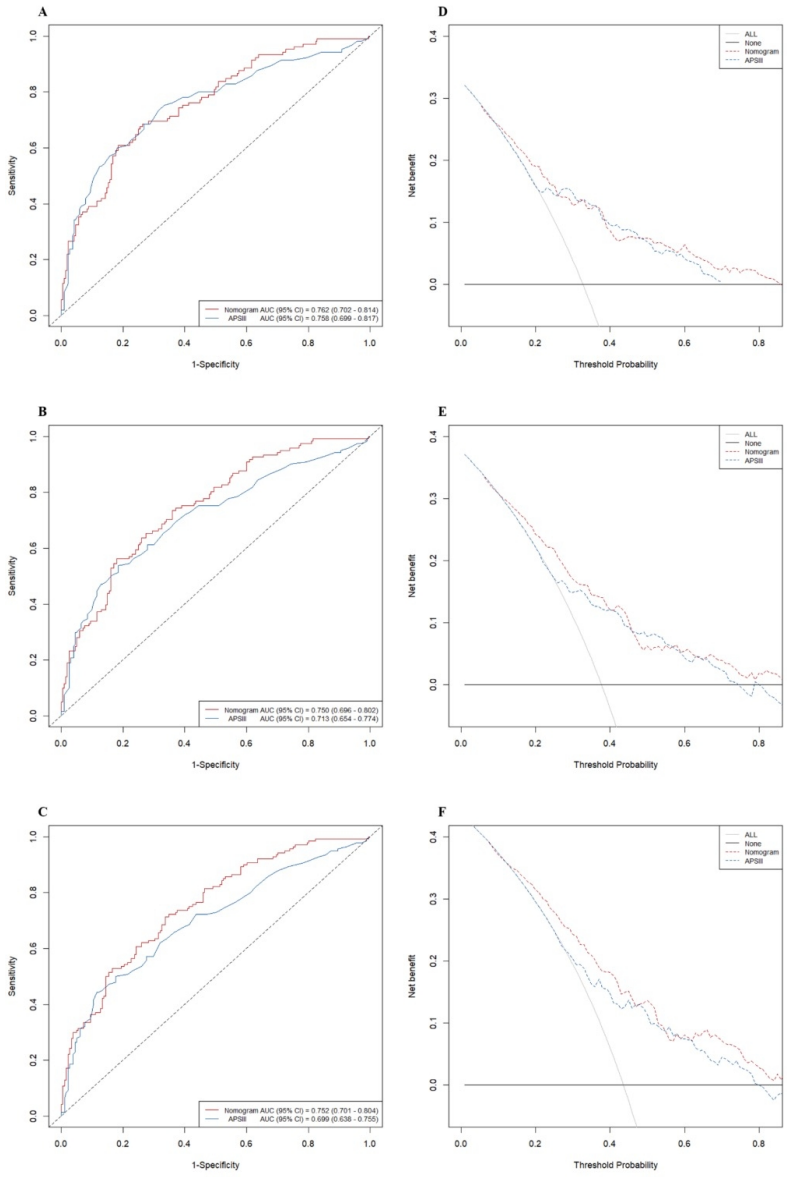


Figure 9. Time-dependent ROC curves using the nomogram and APSIII predict overall survival probabilities within 30-day (A), 90-day (B), and 180-day (C) in the training set. Decision curve analysis of nomogram and APSIII for 30-day (D), 90-day (E), and 180-day (F) survival probabilities in the training set. APSIII: acute physiology score III; AUC: area under the curve; CI: confidence interval; ROC: receiver operating characteristic.

118x177mm (220 x 220 DPI)