Development and Validation of a Predictive Model for Poor Pulmonary Rehabilitation Outcomes in Patients After Radical Lung Cancer Surgery

Ann. Ital. Chir., 2026 97, 3: 460–471
https://doi.org/10.62713/aic.4403

Siyuan Zhao[1,†], Jiayi He[1,†], Huirong Sun[1], Yaping Liu[1], Jingwen Guo[1], Huiying Huang[1], Liu Hu[1]

[1]Department of Thoracic Surgery, The Seventh Affiliated Hospital, Sun Yat-sen University, 518107 Shenzhen, Guangdong, China

AIM: This study aims to develop and validate a nomogram model based on surgical parameters and clinical features for predicting the risk of poor pulmonary rehabilitation outcomes in patients with non-small cell lung cancer (NSCLC) after radical surgery.

METHODS: This retrospective cohort study included 320 patients who underwent radical lung cancer surgery between June 2021 and December 2024. Patients were divided into a development cohort (n = 224) and a validation cohort (n = 96) at a 7:3 ratio. A composite surgical trauma score (STS) was developed to integrate key intraoperative variables. Independent predictors were identified through multivariate logistic regression analysis, including surgical trauma score, operative time, intraoperative blood loss, and other clinical factors. A nomogram model was constructed and validated using the bootstrap resampling (1000 repetitions). Model performance was assessed using the area under the curve (AUC) and calibration metrics.

RESULTS: The incidence of poor pulmonary rehabilitation was 38.4% in the development cohort and 36.5% in the validation cohort. In the development cohort, 11.2% of patients had a forced expiratory volume in 1 second (FEV1) recovery rate $<$80% alone, 15.2% had a modified Medical Research Council (mMRC) score $\geq$2 alone, and 12.0% met both criteria. Multivariable analysis predicted five independent variables: high versus low surgical trauma score (odds ratio (OR) = 3.09, 95% confidence interval (CI): 1.71–5.58), preoperative FEV1% predicted $<$70% (OR = 2.67, 95% CI: 1.53–4.67), operative time $\geq$180 minutes (OR = 2.52, 95% CI: 1.41–4.50), intraoperative blood loss $\geq$80 mL (OR = 2.25, 95% CI: 1.27–3.99), and age $\geq$65 years (OR = 1.92, 95% CI: 1.10–3.35). The nomogram demonstrated good discrimination, with an AUC of 0.84 (95% CI: 0.78–0.90) in the development cohort and 0.79 (95% CI: 0.70–0.88) in the validation cohort. Calibration was satisfactory in both cohorts (p = 0.23 in the development cohort; p = 0.31 in the validation cohort). Decision curve analysis revealed meaningful net benefits across a wide range of threshold probabilities. Predictive performance remained consistent across subgroups defined by age groups, surgical trauma strata, and tumor-node-metastasis (TNM) stages. Furthermore, sensitivity analyses using alternative endpoint definitions and various approaches for handling missing data further confirmed the robustness of the model.

CONCLUSIONS: We developed and validated a predictive nomogram for poor pulmonary rehabilitation outcomes after radical lung cancer surgery. This model incorporates readily available clinical and surgical parameters, providing individualized risk assessment that may facilitate personalized rehabilitation strategies and improve postoperative management.

Keywords: lung cancer; pulmonary rehabilitation; risk prediction; nomogram; poor outcome

Introduction

Lung cancer remains the leading cause of cancer-related mortality worldwide, with non-small cell lung cancer (NSCLC) accounting for approximately 85% of all cases [1]. For patients with early-stage NSCLC, radical surgical resection represents the primary treatment option, offering a greater chance of long-term survival [2,3]. However, 30–40% of patients experience postoperative pulmonary rehabilitation impairment following radical lung cancer surgery, characterized by insufficient recovery of lung function and persistent dyspnea [4]. These limitations significantly compromise patients' quality of life and are also associated with the higher risk of postoperative complications, prolonged hospitalization, and increased healthcare costs [5,6].

Current clinical practice primarily relies on physicians' expertise and limited clinical indicators to predict postoperative pulmonary rehabilitation outcomes [7]. Traditional assessment frameworks, including tumor-node-metastasis (TNM) staging, focus primarily on tumor anatomy and are inadequate for predicting functional recovery [8,9]. Consequently, patients with the same TNM stage may exhibit substantially different rehabilitation trajectories, highlighting that anatomical indicators alone do not adequately reflect individual patient variability [10,11].

Submitted: 17 October 2025 Revised: 8 December 2025 Accepted: 11 December 2025 Published: 10 March 2026
Correspondence to: Liu Hu, Department of Thoracic Surgery, The Seventh Affiliated Hospital, Sun Yat-sen University, 518107 Shenzhen, Guangdong, China (e-mail: huliu5@mail.sysu.edu.cn).
[†] *These authors contributed equally.*
Editor: Marcello Migliore

Evidence indicates that existing predictive models predominantly depend on preoperative, static indicators such as baseline pulmonary function, age, and smoking history [12,13]. While these factors provide certain prognostic information, they do not capture the dynamic physiological changes associated with surgical trauma or postoperative recovery. Furthermore, most commonly adopted assessment tools have been derived from Western cohorts, and their generalizability to Chinese populations remains uncertain due to ethnic differences and variations in healthcare systems and perioperative management [14].

To address these limitations, integrating surgical stress-related parameters (such as operative time and intraoperative blood loss) with baseline characteristics (including age and preoperative pulmonary function) may offer more robust and tailored risk assessment. This comprehensive framework could enhance risk stratification, facilitate earlier intervention for high-risk patients, and support efficient allocation of clinical resources. As individual surgical indicators may not adequately capture the cumulative impact of diverse surgical experiences, we developed a novel composite surgical trauma score (STS). The STS integrates surgical strategy, operative time, and intraoperative blood loss into a single, clinically interpretable variable, offering a more holistic quantification of surgical trauma, which is particularly pertinent in the current era of minimally invasive approaches.

To our knowledge, an integrated score specifically designed to predict pulmonary rehabilitation outcomes after radical lung cancer surgery and validated in a Chinese cohort has not been previously reported. Therefore, this study aimed to develop and validate a nomogram model that incorporates surgical parameters with key clinical features to predict the risk of suboptimal pulmonary rehabilitation outcomes in patients with NSCLC after radical resection. By providing individualized risk assessment, this model seeks to facilitate personalized rehabilitation strategies and improve postoperative clinical management.

Methods

Study Design and Recruitment of the Study Cohort

This retrospective cohort study was conducted in accordance with the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) statement [15]. Consecutive patients who underwent radical lung cancer surgery at The Seventh Affiliated Hospital, Sun Yat-sen University between June 2021 and December 2024 were included. Patients were randomly divided into development (70%) and validation (30%) cohorts using stratified sampling to maintain comparability across key prognostic factors. The 7:3 split was applied to provide adequate data for model construction while preserving an independent dataset for performance assessment. Patients were allocated to training and validation cohorts using stratified sampling based on age, TNM staging, and surgical approach to ensure balanced distribution of key prognostic variables. The study protocol was approved by the Ethics Committee of The Seventh Affiliated Hospital, Sun Yat-sen University (Approval No. KY-2025-474-01), and the requirement for informed consent was waived because the analysis used anonymized retrospective data.

Inclusion criteria for patient enrollment included (1) pathologically confirmed NSCLC, and (2) radical surgical resection conducted at our institution between June 2021 and December 2024. Exclusion criteria were as follows: (1) receipt of preoperative chemotherapy or radiotherapy; (2) coexisting primary lung diseases, such as severe Chronic Obstructive Pulmonary Disease (COPD) or interstitial lung disease; (3) incomplete surgical resection with with positive margins (Residual tumor microscopically/Residual tumor macroscopically) (R1/R2); and (4) patients with missing baseline clinical data or incomplete key outcome data at the 6-month follow-up (forced expiratory volume in 1 second (FEV1) or modified Medical Research Council (mMRC) score).

Data Collection and Variable Definition

Trained investigators retrieve required data from electronic medical records, anesthesia information systems, and pulmonary function databases. Researchers were blinded to the outcome status of included participants. To improve data accuracy and consistency, data were extracted independently by two trained researchers. Discrepancies were resolved through discussion, and when needed, adjudication by a third senior researcher.

Extracted data included demographic characteristics (age, gender, body mass index (BMI)), smoking history ($\geq$10 pack-years), preoperative pulmonary function (FEV1% predicted), Charlson Comorbidity Index (CCI), tumor characteristics (TNM stage based on the American Joint Committee on Cancer (AJCC) 8th edition, pathological subtype). Surgical variables included surgical procedure (video-assisted thoracoscopic surgery (VATS) vs. open thoracotomy), operative time, intraoperative blood loss, extent of resection, and postoperative rehabilitation adherence.

Rehabilitation adherence was defined based on nursing records of daily supervised breathing exercises and ambulation during hospitalization and classified as good ($\geq$80% of activities completed), fair (50–79%), and poor ($<$50%). Operative time was defined as the interval from skin incision to closure, and blood loss was calculated as suction volume plus gauze weight, assuming 1 g equal to 1 mL.

Outcome Definition

The primary outcome was poor pulmonary rehabilitation at 6 months after surgery, defined using a composite endpoint that met either of the following criteria: (1) inadequate lung function recovery, evidenced by an FEV1 recovery rate below 80% calculated as postoperative FEV1/preoperative FEV1 $\times$ 100%, with measurements conducted according

to American Thoracic Society/European Respiratory Society (ATS/ERS) guidelines; or (2) symptomatic dyspnea, defined as a mMRC dyspnea score of 2 or greater, indicating dyspnea when walking faster than peers or the need to stop for breath when walking at a normal pace. The 6-month assessment timepoint was chosen as it represents a crucial phase in postoperative pulmonary recovery, during which significant functional improvements are expected to stabilize.

The FEV1 recovery rate <80% is widely used to indicate insufficient recovery of lung function, reflecting a clinically significant deviation from baseline capacity and a threshold commonly utilized in pulmonary rehabilitation studies. Similarly, an mMRC dyspnea score of 2 or higher is a clinically well-recognized indicator of significant dyspnea, often causing patients to slow down or stop during daily activities, and therefore indicating a compromised functional status and reduced quality of life. All pulmonary function tests were performed by certified technicians using the same calibrated equipment.

Development of the Prediction Model

The prediction model was constructed using multivariate logistic regression. Candidate variables with potential prognostic significance ($p < 0.1$) in univariate analysis were included in the final model. Given the high proportion of VATS (92.8%) and the low median intraoperative blood loss in our cohort, we developed a composite Surgical Trauma Score (STS) to better reflect overall surgical stress. The STS categorized patients as High Trauma if they underwent open thoracotomy, or if they underwent VATS with both operative time ≥180 minutes and intraoperative blood loss ≥80 mL; all other patients were classified as Low Trauma.

To enhance clinical applicability and facilitate interpretation at the point of care, continuous variables (age, operative time, blood loss, and preoperative FEV1% predicted) were dichotomized using established clinical cutoffs or data-driven thresholds. For example, the blood loss of 80 mL or greater corresponded approximately to the 90th percentile in our cohort and was used to define significant intraoperative bleeding in the context of minimally invasive surgery, while age ≥65 years indicated a commonly used geriatric threshold. Although dichotomization can lead to some information loss, this approach was adopted to enhance the utility of the nomogram and to facilitate clear risk stratification.

Preliminary analysis evaluating potential non-linear association with restricted cubic splines did not yield significantly improved model performance or clinical interpretability, supporting the use of a dichotomized approach for this prediction model. This binary classification and other significant variables were entered into a forward stepwise logistic regression model (entry $p = 0.05$; removal $p = 0.10$), which identified five independent predictors. Before multivariate modeling, multicollinearity among independent variables was assessed using variance inflation factors (VIFs); all VIFs were below 5, indicating no significant multicollinearity. A nomogram was then developed based on the final regression coefficients.

Statistical Analysis

Statistical analysis was performed using SPSS 26.0 (IBM Corp., Armonk, NY, USA) and R (v4.2.0, R Foundation for Statistical Computing, Vienna, Austria). Based on the Events Per Variable (EPV) ≥10 principle and an expected event rate of 38%, at least 132 patients were required for model development. The final development cohort included 224 patients, exceeding this requirement. Continuous variables were expressed as mean ± standard deviation (SD) or median (interquartile range [IQR]) (normality-dependent), whereas categorical variables were expressed as counts and percentages (%). Group comparisons were conducted using *t*-test, Mann-Whitney U, χ^2, or Fisher's exact test, as appropriate. The selection of candidate predictors was based on clinical relevance and prior evidence, rather than solely on univariate significance. To avoid overfitting, additional modeling methods, including least absolute shrinkage and selection operator (LASSO), were also applied for sensitivity analysis.

Model performance was internally validated using bootstrap resampling with 1000 repetitions. In each bootstrap sample, the logistic regression model was fitted and evaluated for discrimination (area under the curve (AUC)) and calibration (Hosmer-Lemeshow test and calibration curve). An optimism-corrected AUC was calculated by subtracting the mean optimism across bootstrap samples from the apparent AUC, providing a more reliable estimate of model performance in new patients. Calibration curves were generated to visualize the agreement between predicted risks and observed event rates. Subgroup analyses were conducted by age, surgical trauma score, and TNM stage, and sensitivity analyses were performed using alternative outcome definitions and various approaches to handling missing data.

In addition to the forward stepwise logistic regression, alternative modeling approaches, including penalized logistic regression (LASSO) and Random Forest algorithms, were applied in sensitivity analyses to assess the robustness of predictor selection and overall model performance. A two-tailed *p*-value of less than 0.05 was considered statistically significant.

Results

Comparison of Baseline Characteristics Between the Two Study Groups

A total of 320 patients who underwent radical lung cancer surgery between June 2021 and December 2024 were included and randomly assigned to a development cohort (n = 224) and a validation cohort (n = 96) at a 7:3 ratio. Baseline

Table 1. Baseline characteristics of patients in the development and validation cohorts.

Variable	Overall (n = 320)	Training set (n = 224)	Validation set (n = 96)	*p*-value
Age (years)	62.3 ± 8.5	62.1 ± 8.3	62.6 ± 9.0	0.64
Male [n (%)]	199 (62.2)	138 (61.6)	61 (63.5)	0.74
BMI (kg/m^2)	23.4 ± 3.1	23.2 ± 3.0	23.7 ± 3.4	0.21
Smoking history [n (%)]	156 (48.8)	110 (49.1)	46 (47.9)	0.85
Preoperative FEV1% <70% [n (%)]	88 (27.5)	65 (29.0)	23 (24.0)	0.35
Charlson Comorbidity Index ≥3 [n (%)]	98 (30.6)	69 (30.8)	29 (30.2)	0.92
TNM stage [n (%)]				0.99
Stage 0-I	142 (44.4)	100 (44.6)	42 (43.8)	
Stage II	108 (33.8)	75 (33.5)	33 (34.4)	
Stage III	70 (21.8)	49 (21.9)	21 (21.8)	
Surgical approach [n (%)]				0.96
VATS	297 (92.8)	208 (92.9)	89 (92.7)	
Open thoracotomy	23 (7.2)	16 (7.1)	7 (7.3)	
Operative time ≥180 min [n (%)]	118 (36.9)	83 (37.1)	35 (36.5)	0.92
Intraoperative blood loss ≥80 mL [n (%)]	37 (11.6)	26 (11.6)	11 (11.5)	0.97
Surgical trauma score (high) [n (%)]	54 (16.9)	39 (17.4)	15 (15.6)	0.70
Poor rehabilitation adherence [n (%)]	81 (25.3)	57 (25.4)	24 (25.0)	0.93
Resection extent [n (%)]				0.65
Wedge resection	141 (44.1)	95 (42.4)	46 (47.9)	
Segmentectomy	52 (16.3)	38 (17.0)	14 (14.6)	
Lobectomy or pneumonectomy	127 (39.7)	91 (40.6)	36 (37.5)	

Note: Continuous variables are presented as mean ± standard deviation or median (interquartile range). Categorical variables are presented as a number (percentage). *p*-values were calculated using Student's *t*-test for normally distributed continuous variables, Mann-Whitney U test for non-normally distributed continuous variables, and chi-square test or Fisher's exact test for categorical variables. BMI, body mass index; FEV1, forced expiratory volume in 1 second; VATS, video-assisted thoracoscopic surgery; TNM, tumor-node-metastasis.

demographic and clinical characteristics are summarized in Table 1. There were no statistically significant differences in baseline characteristics between the two cohorts (all $p > 0.05$), indicating good comparability between the training set and the validation set.

Overall, the mean age of the patients was 62.3 ± 8.5 years, with 199 (62.2%) male participants. Consistent with current surgical trends, VATS was the primary surgical approach, conducted in 297 (92.8%). Regarding the extent of resection, sublobar strategies (wedge resection or segmentectomy) were performed in 193 (60.3%) patients, whereas lobectomy or pneumonectomy was performed in 127 (39.7%) patients.

Using the composite surgical trauma score, 54 (16.9%) patients were classified as high trauma, including all patients who underwent open thoracotomy and a subset of VATS patients meeting predefined criteria of prolonged operative time and increased intraoperative blood loss. The distributions of operative time ≥180 minutes (36.9%) and intraoperative blood loss ≥80 mL (11.6%) were comparable between the development and validation cohorts.

Incidence of Poor Pulmonary Rehabilitation Outcomes

The incidence of poor pulmonary rehabilitation outcomes 6 months after surgery was 38.4% (86/224) in the development cohort and 36.5% (35/96) in the validation cohort. The pattern of outcomes was comparable between the two cohorts. In the development cohort, the poor outcome was attributable to insufficient FEV1 recovery alone in 11.2% of patients, clinically symptomatic dyspnea (mMRC ≥2) alone in 15.2%, and the combination of both criteria in 12.0% (Table 2). The similar overall incidence and proportional distribution of each component support a balanced event distribution, which is crucial for robust model validation.

Detailed distributions of the two outcome measures are presented in Table 3. The median FEV1 recovery rate was 85.0% in the development cohort and 84.5% in the validation cohort. Specifically, most patients achieved an FEV1 recovery rate ≥80% (development: 76.8%; validation: 79.2%), while 23.2% and 20.8% had recovery rates below 80%, respectively. Regarding dyspnea, 72.8% of patients in the development cohort reported minimal symptoms (mMRC score 0–1), whereas 27.2% reported more significant exertional dyspnea (mMRC score ≥2).

Predictors of Poor Pulmonary Rehabilitation Outcomes

Univariate analysis identified several potential predictors associated with poor pulmonary rehabilitation outcomes ($p < 0.1$), including age ≥65 years, smoking history, preoper-

Table 2. Incidence and pattern of poor pulmonary rehabilitation outcomes.

Outcome criteria	Development cohort (n = 224)	Validation cohort (n = 96)
Composite endpoint	86 (38.4%)	35 (36.5%)
FEV1 recovery rate <80% only	25 (11.2%)	10 (10.4%)
mMRC score ≥2 only	34 (15.2%)	15 (15.6%)
Both criteria	27 (12.0%)	10 (10.4%)

Note: mMRC, modified Medical Research Council.

Table 3. Distribution of postoperative FEV1 recovery rate and mMRC scores at 6 months.

Variable	Development cohort (n = 224)	Validation cohort (n = 96)
FEV1 recovery rate		
Mean ± SD (%)	82.4 ± 15.2	81.9 ± 15.7
Median (IQR) (%)	85.0 (73.0–94.0)	84.5 (72.0–93.5)
FEV1 recovery rate status [n (%)]		
<80% (Poor recovery)	52 (23.2%)	20 (20.8%)
≥80% (Good recovery)	172 (76.8%)	76 (79.2%)
mMRC score distribution		
Score 0	98 (43.8%)	41 (42.7%)
Score 1	65 (29.0%)	29 (30.2%)
Score 2	42 (18.8%)	17 (17.7%)
Score 3	15 (6.7%)	7 (7.3%)
Score 4	4 (1.8%)	2 (2.0%)

Note: SD, standard deviation; IQR, interquartile range.

Continuous variables are presented as both mean ± standard deviation and median (interquartile range) to better describe the distribution and potential skewness of the data.

ative FEV1% predicted <70%, a Charlson Comorbidity Index ≥3, operative time ≥180 minutes, intraoperative blood loss ≥80 mL and a high surgical trauma score (Table 4).

Multivariate logistic regression analysis with forward stepwise selection identified five independent predictors of poor pulmonary rehabilitation (Table 5). The STS emerged as the strongest predictor, with high surgical trauma associated with a 3.09-fold increased risk of poor outcomes compared to low surgical trauma (odds ratio (OR) = 3.09, 95% confidence interval (CI): 1.71–5.58, $p < 0.001$). Other significant indicators included preoperative FEV1% predicted <70% (OR = 2.67, 95% CI: 1.53–4.67, $p = 0.001$), operative time ≥180 minutes (OR = 2.52, 95% CI: 1.41–4.50, $p = 0.002$), age ≥65 years (OR = 1.92, 95% CI: 1.10–3.35, $p = 0.022$), and intraoperative blood loss ≥80 mL (OR = 2.25, 95% CI: 1.27–3.99, $p = 0.005$). These findings underscore the multifactorial nature of postoperative pulmonary rehabilitation outcomes, involving both surgical-related factors and patient baseline characteristics.

Development and Validation of the Nomogram Model

A nomogram was developed using the five independent predictors identified in multivariable regression analysis: STS, preoperative FEV1% predicted <70%, operative time ≥180 minutes, intraoperative blood loss ≥80 mL, and age ≥65 years (Fig. 1). Each predictor was weighted based on its corresponding regression coefficient, and the total score was translated into predicted probabilities of pulmonary rehabilitation.

The nomogram demonstrated strong discriminatory performance in the development cohort, yielding an AUC of 0.84 (95% CI: 0.78–0.90). In the validation cohort, the AUC was 0.79 (95% CI: 0.70–0.88), indicating good generalizability (Fig. 2). Calibration curves showed close agreement between predicted probabilities and observed outcomes in both cohorts (Hosmer-Lemeshow test, $p = 0.23$ in development cohort; $p = 0.31$ in validation cohort) (Fig. 3). The calibration slope was 0.98 (95% CI: 0.89–1.07) for the development cohort and 0.95 (95% CI: 0.83–1.07) for the validation cohort, further indicating minimal miscalibration and no significant over- or under-prediction or estimation.

Decision curve analysis demonstrated net clinical benefit of the nomogram across a range of threshold probabilities, particularly between 10% and 60%, in both cohortse (Fig. 4).

Subgroup Analysis and Sensitivity Analysis

Subgroup analyses demonstrated that the nomogram maintained stable predictive performance across various patient strata. The model's discriminatory performance remained good in patients aged <65 vs. ≥65 years (AUC: 0.82 vs. 0.84), in low vs. high surgical trauma groups (AUC: 0.82 vs. 0.85), and across TNM stages 0–I vs. II–III (AUC: 0.81 vs. 0.78). Notably, interaction testing showed no significant differential performance across these subgroups (all p-values for interaction >0.05), confirming the broad applicability of the model (Table 6). Furthermore, we conducted sensitivity analyses to assess the model's robustness. The model's performance remained strong under alterna-

Table 4. Univariate analysis of poor pulmonary rehabilitation outcomes.

Variable	Good rehabilitation (n = 138)	Poor rehabilitation (n = 86)	χ^2/t	*p*-value
Age ≥65 years [n (%)]	42 (30.4)	45 (52.3)	10.69	0.001
Male [n (%)]	82 (59.4)	56 (65.1)	0.73	0.394
BMI (kg/m^2)	23.3 ± 3.0	23.1 ± 3.4	1.42	0.157
Smoking history [n (%)]	61 (44.2)	49 (57.0)	3.46	0.063
Preoperative FEV1% <70% [n (%)]	31 (59.4)	34 (39.5)	7.50	0.006
Charlson Comorbidity Index ≥3 [n (%)]	35 (25.4)	34 (39.5)	4.99	0.026
Operative time ≥180 min [n (%)]	41 (29.7)	42 (48.8)	8.31	0.004
Intraoperative blood loss ≥80 mL [n (%)]	10 (7.2)	16 (18.6)	6.66	0.010
Surgical trauma score (high) [n (%)]	15 (10.9)	24 (27.9)	10.70	0.001

Note: Continuous variables are presented as mean ± standard deviation. Categorical variables are presented as a number (percentage). *p*-values were calculated using Student's *t*-test for continuous variables and chi-square test for categorical variables.

Table 5. Independent predictors of poor pulmonary rehabilitation.

Variable	β-coefficient	SE	Wald χ^2	*p*-value	OR	95% CI
Surgical trauma score (high vs. low)	1.128	0.301	14.05	<0.001	3.09	1.71–5.58
Preoperative FEV1% <70%	0.982	0.285	11.87	0.001	2.67	1.53–4.67
Operative time ≥180 min	0.924	0.295	9.81	0.002	2.52	1.41–4.50
Intraoperative blood loss ≥80 mL	0.812	0.291	7.79	0.005	2.25	1.27–3.99
Age ≥65 years	0.652	0.284	5.28	0.022	1.92	1.10–3.35
Intercept	–2.943	0.425	47.98	<0.001	—	—

Note: OR, odds ratio; CI, confidence interval; SE, standard error. Variables were selected using forward stepwise logistic regression.

Table 6. Model discriminatory performance across different subgroups.

Subgroup	Sample size	Events	AUC (95% CI)	*p*-value (Interaction)
Age <65 years	137	49	0.82 (0.74–0.89)	0.62
Age ≥65 years	87	36	0.84 (0.75–0.93)	
Low surgical trauma score	185	72	0.82 (0.75–0.89)	0.57
High surgical trauma score	39	24	0.85 (0.73–0.97)	
TNM stage 0–I	100	35	0.81 (0.74–0.88)	0.67
TNM stage II-III	124	51	0.78 (0.66–0.89)	

Note: *p*-values for interaction were calculated using likelihood ratio tests.

tive outcome definitions; when FEV1 recovery rate <80% or mMRC score ≥2 was used as the standard endpoints, the AUCs in the development cohort were 0.82 and 0.80, respectively. The model's integrity was confirmed, and consistent results were found with different approaches to missing data (multiple imputation vs. complete-case analysis) and with alternative modeling approaches (LASSO regression vs. random forest). In all these scenarios, discrimination remained in a clinically useful range, with AUCs consistently between 0.80 and 0.84 in the development cohort and 0.77 and 0.79 in the validation cohort, underscoring the robustness of our primary findings (Table 7).

Discussion

In this study, we developed and internally validated a nomogram model to predict poor pulmonary rehabilitation outcomes after radical lung cancer surgery. The model incorporated five readily available predictors: a novel composite STS, preoperative FEV1% predicted <70%, operative time ≥180 minutes, intraoperative blood loss ≥80 mL, and age ≥65 years. The nomogram demonstrated good discrimination and calibration in both development and validation cohorts, suggesting its potential clinical utility for individualized risk stratification in postoperative care.

Our findings highlight the significance of integrating an overall measure of surgical stress with patient baseline characteristics. The newly constructed STS emerged as the strongest predictor in the model. This composite score offers a more nuanced assessment of surgical burden than a simple classification by surgical approach (VATS vs. open). By integrating operative approach, operative time, and intraoperative blood loss, the STS also alleviates the statistical challenge created by the relatively low incidence of open thoracotomy in modern practice. More importantly, within our cohort, this scoring system can identify high-risk patients even in a predominantly VATS-treated population,

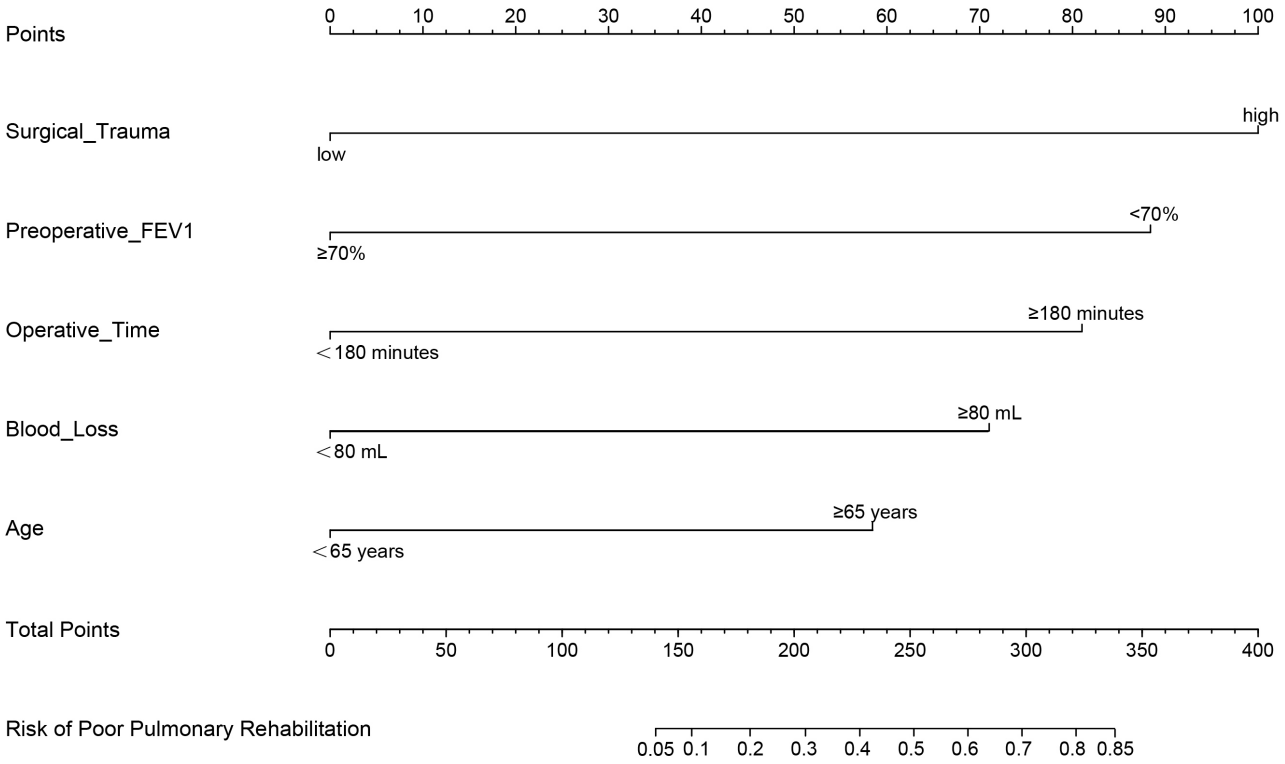

Fig. 1. Nomogram for predicting the risk of poor pulmonary rehabilitation outcomes at 6 months after radical lung cancer surgery. Note: In the nomogram, Surgical Trauma refers to the surgical trauma score (STS) categorized as high versus low. Preoperative FEV1 refers to preoperative forced expiratory volume in 1 second percent predicted (<70% vs. ≥70%). Operative Time and Blood Loss are defined as ≥180 minutes and ≥80 mL, respectively. Age is categorized as ≥65 years. All predictors and their cutoff values are consistent with those presented in Table 5.

Table 7. Sensitivity analysis using different outcome definitions.

	Development cohort AUC (95% CI)	Validation cohort AUC (95% CI)
Outcome Definition		
Primary composite endpoint	0.84 (0.78–0.90)	0.79 (0.71–0.87)
FEV1 recovery rate <80% only	0.82 (0.76–0.88)	0.78 (0.69–0.86)
mMRC score ≥2 only	0.80 (0.73–0.87)	0.77 (0.68–0.85)
Both FEV1 <80% and mMRC ≥2	0.83 (0.77–0.90)	0.79 (0.70–0.88)
Missing data handling		
Multiple imputation	0.83 (0.77–0.89)	0.78 (0.70–0.86)
Complete case analysis	0.81 (0.74–0.88)	0.78 (0.69–0.86)
Alternative modeling approaches		
Logistic regression with LASSO	0.81 (0.75–0.87)	0.78 (0.69–0.87)
Random forest	0.80 (0.74–0.86)	0.77 (0.68–0.86)

Note: LASSO, least absolute shrinkage and selection operator.

particularly among those undergoing longer and more complex procedures [16–18]. These observations support the concept that surgical trauma is a continuum, and the STS provides a practical tool to quantify that continuum in routine clinical care.

While composite scoring systems are commonly used for surgical risk assessment in other settings, like EuroSCORE in cardiac surgery and APACHE II in critical care during general surgery, to our knowledge, a similar integrated score specifically designed to predict pulmonary rehabilitation outcomes after radical lung cancer surgery and validated within a Chinese population appears to be novel. The key strength of STS is its specific focus on aspects of surgical trauma, which is most likely to affect postoperative

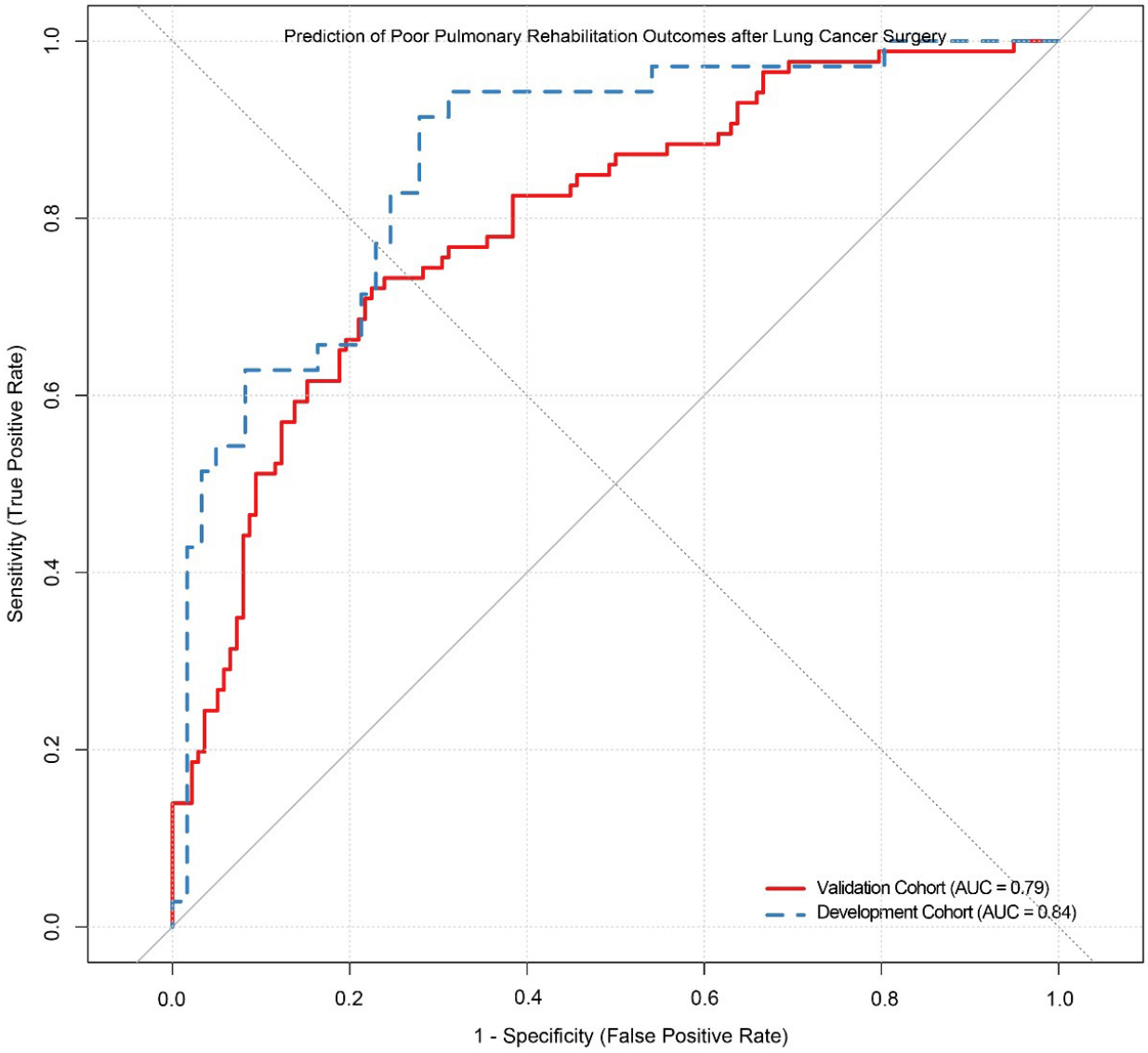

Fig. 2. Evaluation of model discriminatory performance (ROC curve). ROC, Receiver Operating Characteristic; AUC, area under the curve.

pulmonary recovery in thoracic surgery, integrating factors that are readily available and clinically relevant for this patient group.

Operative time ≥180 minutes and intraoperative blood loss ≥80 mL were also identified as independent predictors, underscoring their distinct contributions to the overall burden of surgical trauma. Prolonged operative time or procedures and increased blood loss indicate greater intraoperative complexity and can lead to more extensive tissue injury, a severe inflammatory response, and a higher risk of perioperative complications, all of which may adversely affect postoperative pulmonary recovery [19–21]. Notably, the blood loss threshold of ≥80 mL, corresponding to approximately the 90th percentile in our cohort, is more sensitive for identifying clinically relevant bleeding in the current minimally invasive era than traditional benchmarks such as 200 mL. This demonstrates that even a moderate but statistically significant increase in intraoperative bleeding remains a critical early warning sign of impaired postoperative recovery.

Age ≥65 years and preoperative FEV1% predicted <70% were identified as patient-related factors, not significantly different from previous studies [22–26]. Elderly patients often have reduced physiological reserve and diminished adaptive capacity to surgical stress, whereas impaired preoperative pulmonary function indicates limited ability to compensate for postoperative respiratory demands [27–29]. These factors should be carefully considered during preoperative assessment and perioperative management planning. The mean age in our cohort was 62.3 years, which

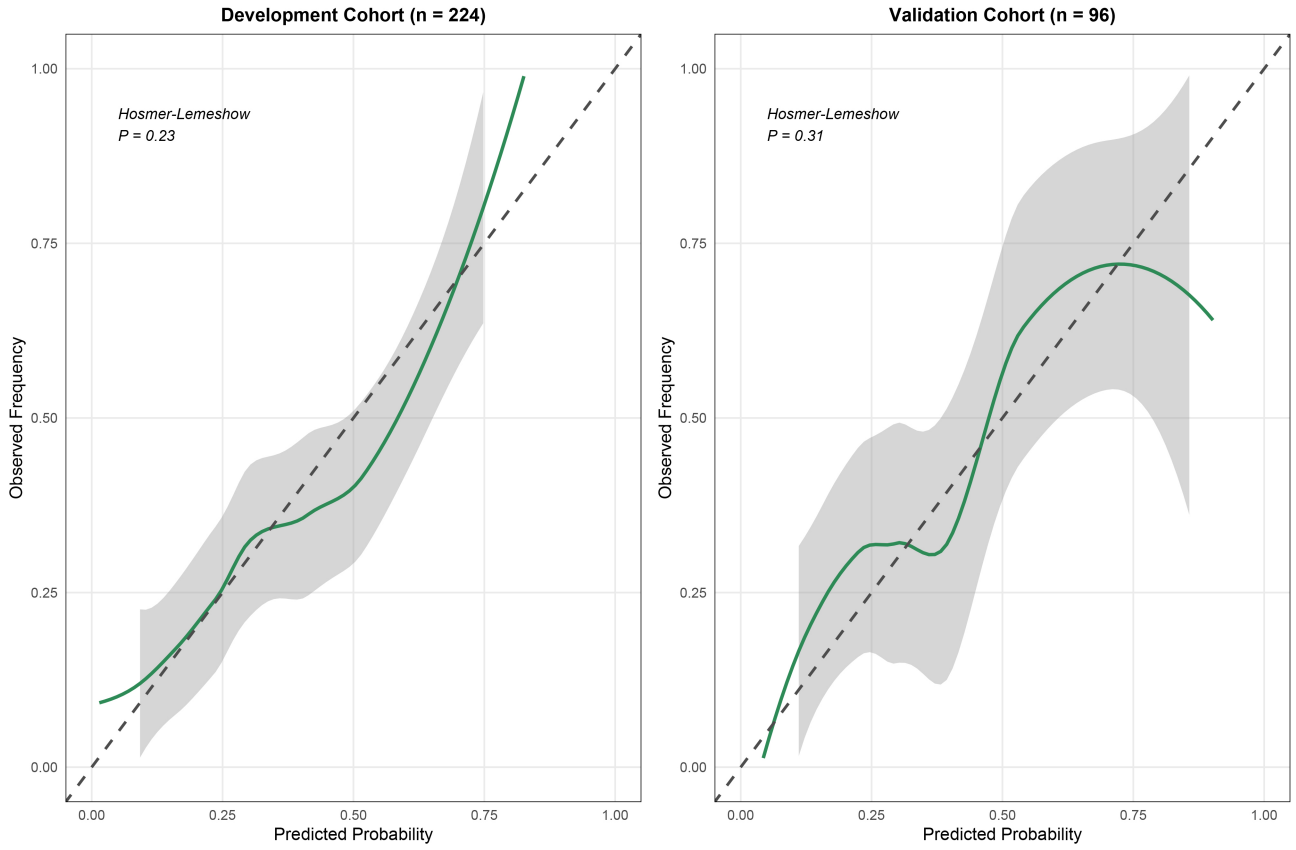

Fig. 3. Model calibration evaluation.

reflects the typical demographic of the lung cancer cohort. Only 27.5% of patients had a preoperative FEV1% predicted below 70%, indicating that most patients had relatively preserved lung function before surgery.

The component-specific pattern of poor rehabilitation provides additional insights into postoperative impairment. The observation that 15.2% of patients experienced clinically significant dyspnea despite FEV1 recovery suggests factors beyond lung function, such as physical deconditioning, pain, or altered breathing mechanics, may contribute to perceived dyspnea. Conversely, 11.2% demonstrated inadequate FEV1 recovery in the absence of significant dyspnea, highlighting that lung function impairment may not always align with symptom burden. Overall, these results support the need to integrate both objective pulmonary function symptoms and subjective patient-related measures in a comprehensive assessment of postoperative rehabilitation.

The nomogram model developed in this study offers several advantages over commonly used assessment tools. Its key strength lies in the integration of a composite STS with established patient-related factors, providing a more holistic and accurate risk assessment than models relying on isolated preoperative or intraoperative variables. Additionally, the use of a composite endpoint that combines objective FEV1 decline with the subjective patient-reported mMRC dyspnea score better captures the multifaceted nature of poor pulmonary rehabilitation, reflecting both physiological limitation and its impact on quality of life. This model offers a clinically relevant and comprehensive assessment: all variables are readily available in standard perioperative care, enhancing feasibility for widespread real-world implementation. The nomogram provides a clear, visual and intuitive tool for individualized risk assessment and may facilitate shared decision-making and communication between healthcare providers and patients. The model also maintained consistent performance across different subgroups, including age categories, surgical trauma strata, and TNM stages, indicating its broad applicability.

The clinical implications of this predictive model are meaningful. Patients categorized as high-risk patients by the nomogram could be prioritized for early, targeted postoperative approaches, such as more intensive pulmonary rehabilitation, optimized pain management, and close monitoring for complications. Conversely, low-risk patients may be managed with less intensive follow-up, which could reduce healthcare costs and improve resource allocation. This risk-stratified approach aligns with the current trend toward personalized perioperative management in thoracic surgery [30]. The decision curve analysis further demonstrated that the model provides net clinical benefits across a wide range of threshold probabilities, supporting its clinical utility.

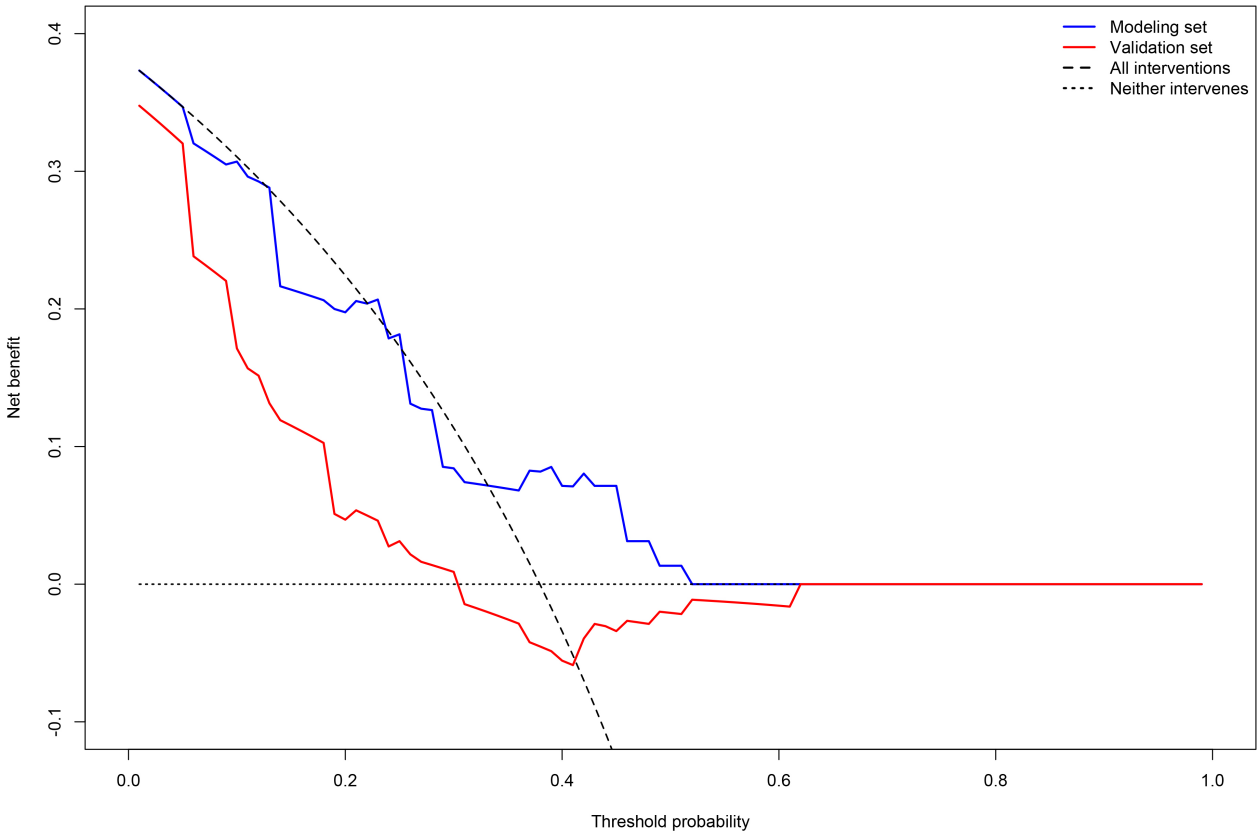

Fig. 4. Decision curve analysis of the nomogram.

Several limitations of this study should be acknowledged when implementing these findings. First, the retrospective design inherently carries risks of selection bias and information bias, despite rigorous data extraction and quality control measures. Although data on postoperative rehabilitation adherence were recorded, reliance on medical record documentation may not fully capture dynamic and fluctuating patterns of patient participation. Prospective monitoring approaches, such as wearable devices or structured daily logs, would provide a more accurate measure of adherence and its association with outcomes.

Second, this was a single-institution study conducted in China, which may limit the generalizability of our findings to diverse patient populations and healthcare settings, particularly in Western countries, due to potential ethnic differences, surgical practices, perioperative care, and rehabilitation protocols. The internal validation cohort was also relatively small (n = 96). Furthermore, the dichotomization of continuous variables, while enhancing clinical adoptability, may reduce precision and may not translate optimally to populations in which different thresholds are more appropriate. Consequently, external validation in independent, multi-center cohorts is an imperative next step to establish broader applicability.

Third, although the model was internally validated using bootstrap resampling (which helps correct for optimism), this approach cannot replace true external validation. Fourth, the scope of predictors included in our model has inherent limitations. We did not incorporate genetic or molecular markers that could add additional prognostic information, and preoperative pulmonary function was represented only by FEV1% predicted. Incorporating additional pulmonary function indicators such as FVC% and DLCO%, which were not consistently available in this retrospective dataset, may improve predictive performance and clinical interpretability.

Finally, outcomes were evaluated only up to 6 months after surgery, and longer-term outcomes remain undetermined. Future research should therefore focus on multicenter prospective external validation, incorporation of novel biomarkers, and evaluation of whether model-guided risk stratification enhances clinical decision-making and patient outcomes. Developing a simplified point-based risk scoring system from the nomogram may further improve its practical utility in routine clinical settings, and integrating the model within an electronic health record system could facilitate real-time decision support at the point of care.

Conclusions

In conclusion, we developed and validated a nomogram model to predict poor pulmonary rehabilitation outcomes in patients after radical lung cancer surgery. The model incorporates five routinely available predictors: the composite STS, preoperative FEV1% predicted $<70\%$, operative time ≥ 180 minutes, intraoperative blood loss ≥ 80 mL, and age ≥ 65 years. The nomogram demonstrated good discrimination and calibration in both development and validation cohorts, suggesting its potential clinical utility for individualized risk assessment and early, targeted postoperative intervention. By facilitating risk-adapted follow-up and rehabilitation planning, this tool may help optimize postoperative management and improve patient prognosis after lung cancer surgery.

Availability of Data and Materials

The data analyzed are available from the corresponding authors upon reasonable request.

Author Contributions

SYZ and JYH were primarily responsible for data collection and analysis, and contributed to the draft of the manuscript. HRS and YPL contributed to data analysis and interpretation. JWG and HYH contributed to the study's methodology and analysis and participated in writing. LH contributed to the project's overall management and participated in the study's design and implementation. All authors contributed to the critical revision of the manuscript for important intellectual content. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

The study was approved by the Ethics Committee of The Seventh Affiliated Hospital, Sun Yat-sen University (Approval No. KY-2025-474-01), and the requirement for informed consent was waived because the analysis used anonymized retrospective data. The study conformed to the provisions of the Declaration of Helsinki.

Acknowledgment

Not applicable.

Funding

This research received no external funding.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Rahal Z, El Darzi R, Moghaddam SJ, Cascone T, Kadara H. Tumour and microenvironment crosstalk in NSCLC progression and response to therapy. Nature Reviews. Clinical Oncology. 2025; 22: 463–482. https://doi.org/10.1038/s41571-025-01021-1.

[2] Jeon H, Wang S, Song J, Gill H, Cheng H. Update 2025: Management of Non Small-Cell Lung Cancer. Lung. 2025; 203: 53. https://doi.org/10.1007/s00408-025-00801-x.

[3] Wang M, Herbst RS, Boshoff C. Toward personalized treatment approaches for non-small-cell lung cancer. Nature Medicine. 2021; 27: 1345–1356. https://doi.org/10.1038/s41591-021-01450-2.

[4] Tenconi S, Mainini C, Rapicetta C, Braglia L, Galeone C, Cavuto S, *et al*. Rehabilitation for lung cancer patients undergoing surgery: results of the PUREAIR randomized trial. European Journal of Physical and Rehabilitation Medicine. 2021; 57: 1002–1011. https://doi.org/10.23736/S1973-9087.21.06789-7.

[5] Klimczak M, Piekielny D, Antczak A, Śmigielski J, Tworek D. Effectiveness of pulmonary rehabilitation in patients with chronic obstructive pulmonary disease after lobectomy due to non-small cell lung cancer - a single-center retrospective study. Advances in Respiratory Medicine. 2021; 89: 247–253. https://doi.org/10.5603/ARM.a2021.0060.

[6] Iacobescu R, Antoniu SA, Popa AD, Pavel-Tanase M, Stratulat TA. Preoperative frailty screening in elderly patients with non-small cell lung cancer surgery: an essential step for a good surgical outcome. Expert Review of Respiratory Medicine. 2024; 18: 99–110. https://doi.org/10.1080/17476348.2024.2349579.

[7] Haesevoets S, Arents E, Cops D, Quadflieg K, Criel M, Ruttens D, *et al*. The impact of lung surgery, with or without (neo-)adjuvant therapy, on physical functioning in patients with nonsmall cell lung cancer: a scoping review. European Respiratory Review: an Official Journal of the European Respiratory Society. 2025; 34: 240156. https://doi.org/10.1183/16000617.0156-2024.

[8] Klug M, Kirshenboim Z, Truong MT, Sorin V, Ofek E, Agrawal R, *et al*. Proposed Ninth Edition TNM Staging System for Lung Cancer: Guide for Radiologists. Radiographics: a Review Publication of the Radiological Society of North America, Inc. 2024; 44: e240057. https://doi.org/10.1148/rg.240057.

[9] Lundström S, Agger E, Lydrup ML, Jörgren F, Buchwald P. Rethinking TNM: Tumor deposit-based prognostic models may improve N-staging in colorectal cancer. European Journal of Surgical Oncology: the Journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology. 2025; 51: 110420. https://doi.org/10.1016/j.ejso.2025.110420.

[10] Safi S, Gysan MR, Weber D, Behnisch R, Muley T, Allgäuer M, *et al*. Peri- and postoperative morbidity and mortality in older patients with non-small cell lung cancer: a matched-pair study. World Journal of Surgical Oncology. 2024; 22: 213. https://doi.org/10.1186/s12957-024-03491-6.

[11] Zhang K, Liu W, Zhao Y, Gao X, Dai W, Zhou X, *et al*. Comparison of early postoperative patient-reported outcomes after multiportal robotic-assisted thoracoscopic surgery and uniportal video-assisted thoracoscopic surgery for non-small cell lung cancer. European Journal of Surgical Oncology: the Journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology. 2024; 50: 108481. https://doi.org/10.1016/j.ejso.2024.108481.

[12] Vedire Y, Nitsche L, Tiadjeri M, McCutcheon V, Hall J, Barbi J, *et al*. Skeletal muscle index is associated with long term outcomes after lobectomy for non-small cell lung cancer. BMC Cancer. 2023; 23: 778. https://doi.org/10.1186/s12885-023-11210-9.

[13] Yang R, Li W, Yu S, Wu Z, Zhang H, Liu X, *et al*. Enhanced NSCLC subtyping and staging through attention-augmented multi-task deep learning: A novel diagnostic tool. International Journal of Medical Informatics. 2025; 193: 105694. https://doi.org/10.1016/j.ijmedinf.2024.105694.

[14] Presley CJ, Tang J, Benedict J, Grogan M, Reisinger S, Janse S, *et al*. Functional resilience and overall survival in adults treated for advanced non-small-cell lung cancer. Lung Cancer (Amsterdam,

Netherlands). 2024; 196: 107953. https://doi.org/10.1016/j.lungcan.2024.107953.

[15] Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooft L, Reitsma JB, *et al*. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open. 2021; 11: e048008. https://doi.org/10.1136/bmjopen-2020-048008.

[16] Zhou Y, Yu T, Zhang Y, Qian L, Xia Q. Comparison of surgical outcomes and prognosis between wedge resection and simple Segmentectomy for GGO diameter between 2 cm and 3 cm in non-small cell lung cancer: a multicenter and propensity score matching analysis. BMC Cancer. 2022; 22: 71. https://doi.org/10.1186/s12885-021-09129-0.

[17] Azari F, Kennedy GT, Bou-Samra P, Singhal S. Intraoperative molecular imaging in thoracic oncology: pushing the boundaries of precision resection for occult non-small cell lung cancer in the era of minimally invasive surgery. Journal of Thoracic Disease. 2022; 14: 3688–3691. https://doi.org/10.21037/jtd-22-1100.

[18] Geropoulos G, Esagian SM, Skarentzos K, Ziogas IA, Katsaros I, Kosmidis D, *et al*. Video-assisted thoracoscopic versus open sleeve lobectomy for non-small cell lung cancer: A systematic review and meta-analysis from six comparative studies. Asian Cardiovascular & Thoracic Annals. 2022; 30: 881–893. https://doi.org/10.1177/02184923221115970.

[19] Zhang W, Chen S, Lin X, Chen H, He R. Lobectomy versus segmentectomy for stage IA3 (T1cN0M0) non-small cell lung cancer: a meta-analysis and systematic review. Frontiers in Oncology. 2023; 13: 1270030. https://doi.org/10.3389/fonc.2023.1270030.

[20] Zhang W, Zhu D. Therapeutic effects of single-port thoracoscopic anatomical segmentectomy on early-stage non-small-cell lung cancer. Journal of Minimal Access Surgery. 2025; 21: 19–24. https://doi.org/10.4103/jmas.jmas_316_23.

[21] Matsubara T, Yamaguchi M, Shimokawa M, Okamoto I. Phase II Trial of Adjuvant Atezolizumab Therapy in Elderly Patients with Completely Resected Stage II/III Non-Small Cell Lung Cancer: RELIANCE Trial. Clinical Lung Cancer. 2024; 25: 280–283. https://doi.org/10.1016/j.cllc.2024.01.009.

[22] Park JE, Hong KS, Choi SH, Lee SY, Shin KC, Jang JG, *et al*. Durvalumab Consolidation After Chemoradiotherapy in Elderly Patients With Unresectable Stage III NSCLC: A Real-World Multicenter Study. Clinical Lung Cancer. 2024; 25: 354–364. https://doi.org/10.1016/j.cllc.2024.02.006.

[23] Baladi A, Tafenzi HA, Zouiten O, Afani L, Essaadi I, El Fadli M, *et al*. Immunotherapy for Elderly Patients with Advanced Non-Small Cell Lung Cancer: Challenges and Perspectives. International Journal of Molecular Sciences. 2025; 26: 2120. https://doi.org/10.3390/ijms26052120.

[24] Spaans LN, Dijkgraaf MGW, Susa D, de Loos ER, Mourisse JMJ, Bouwman RA, *et al*. Intercostal or Paravertebral Block vs Thoracic Epidural in Lung Surgery: A Randomized Noninferiority Trial. JAMA Surgery. 2025; 160: 855–864. https://doi.org/10.1001/jamasurg.2025.1899.

[25] Casaluce F, Gridelli C. Combined chemo-immunotherapy in advanced non-small cell lung cancer: feasible in the elderly? Expert Opinion on Emerging Drugs. 2023; 28: 121–127. https://doi.org/10.1080/14728214.2023.2211346.

[26] Wu GJ, Zhang Y, Liang R, Peng L, Zhang SH. Comparison of Survival Outcomes of Early-Stage Non-Small-Cell Lung Cancer in Elderly Patients (≥70 years) Treated With Stereotactic Body Radiotherapy Versus Surgical Resection. World Journal of Surgery. 2025; 49: 1160–1171. https://doi.org/10.1002/wjs.12584.

[27] Lindqvist J, Jekunen A, Sihvo E, Johansson M, Andersén H. Effect of adherence to treatment guidelines on overall survival in elderly non-small-cell lung cancer patients. Lung Cancer (Amsterdam, Netherlands). 2022; 171: 9–17. https://doi.org/10.1016/j.lungcan.2022.07.006.

[28] Shi P, Li Z, Zhang Y, Shen C, Xia Q, Cao C, *et al*. Surgery or radiotherapy improves survival in elderly patients with early non-small cell lung cancer: A population-based analysis. Journal of Cancer Research and Therapeutics. 2024; 20: 1251–1257. https://doi.org/10.4103/jcrt.jcrt_973_23.

[29] Xu Z, Liu C, Zhu Y, Zou Z, Xie T, Xing P, *et al*. Efficacy of first-line treatments in the elderly and non-elderly patients with advanced epidermal growth factor receptor mutated, non-small cell lung cancer: a network meta-analysis. BMC Cancer. 2022; 22: 514. https://doi.org/10.1186/s12885-022-09592-3.

[30] Liu J, Huang X, Yang Y, Lv W, Wang Y, Xia P, *et al*. Comparison of efficacy and safety of neoadjuvant immunochemotherapy in young and elderly patients with IIA-IIIB non-small-cell lung cancer in real-world practice. BMC Pulmonary Medicine. 2024; 24: 592. https://doi.org/10.1186/s12890-024-03417-8.