British Journal of Hospital Medicine

Br. J. Hosp. Med. (Lond) **2026**; 87(8): 57260
https://doi.org/10.31083/BJHM57260

Article

Development of a LASSO Regression-Based Nomogram to Predict Post-Neoadjuvant Chemotherapy Infection Risk in Lung Cancer Patients

Lijie Zhou[1], Xuejia Chen[1], Juanli Zeng[1], Jun Cao[1], Jiangchuan Chen[1], Jianmin Li[1,*], Huan Yang[1,*]

[1]Department of Pulmonary and Critical Care Medicine, Hunan Provincial People's Hospital, The First Affiliated Hospital of Hunan Normal University, 410005 Changsha, Hunan, China

*Correspondence: Leo780307@sina.com (Jianmin Li); huany2020@126.com (Huan Yang)

Academic Editor: John Alcolado

Submitted: 3 July 2025 Revised: 10 December 2025 Accepted: 16 December 2025 Published: 24 August 2026

Abstract

Aims/Background: Post-chemotherapy infection is a major cause of treatment failure in lung cancer (LC) patients undergoing neoadjuvant chemotherapy (NAC). This study aimed to identify risk factors for post-NAC infection using the Least Absolute Shrinkage and Selection Operator (LASSO) regression and to develop and validate a corresponding risk prediction model. **Methods**: Clinical data from 144 LC patients who underwent NAC at Hunan Provincial People's Hospital between January 2021 and December 2024 were retrospectively analysed. Patients were stratified into a non-infection group (n = 81) and an infection group (n = 63) based on the occurrence of infection. LASSO-logistic regression was used to identify risk factors for post-chemotherapy infection. A risk prediction nomogram was subsequently constructed and validated based on these factors. **Results**: The optimal LASSO model was selected at lambda.1se (λ = 0.074), which retained 9 of the 15 candidate predictors. Eastern Cooperative Oncology Group Performance Status (ECOG-PS) $\geq$2, recent invasive procedures/catheterization, and Nutritional Risk Screening 2002 (NRS-2002) score $\geq$3 were identified as significant risk factors for post-chemotherapy infection, while a high cluster of differentiation 4-positive ($CD4^+$) count served as a protective factor ($p < 0.05$). A nomogram incorporating these variables was subsequently developed. Internal validation using the bootstrap method (1000 iterations) demonstrated a good predictive performance with an area under the curve (AUC) of 0.831 (95% confidence interval [CI]: 0.761–0.901, $p < 0.001$), sensitivity of 76.2%, and specificity of 79.0% at the optimal cutoff. The calibration curve demonstrated good agreement between predicted and observed outcomes. Decision curve analysis (DCA) confirmed the clinical utility of the nomogram by showing a positive net benefit across a wide range of threshold probabilities. **Conclusion**: The LASSO-derived nomogram integrates key variables (ECOG-PS $\geq$2, invasive procedures/catheterization, NRS-2002 score $\geq$3, and $CD4^+$ level) to enable individualised prediction of post-NAC infection risk in LC patients. Targeted interventions addressing these risk factors, together with maintenance of protective factors, may help reduce infection rates and improve treatment outcomes, providing a scientific basis for infection prevention and management.

Keywords: lung neoplasms; infection; nomograms; logistic models; risk assessment; prognosis

1. Introduction

Lung cancer (LC) is a highly heterogeneous malignancy that primarily arises from the malignant transformation of bronchial mucosal or alveolar epithelial cells and remains among the top three malignancies worldwide in both incidence and mortality [1,2]. Common clinical manifestations, including cough, chest pain, and dyspnea, severely impair the quality of life and overall survival of patients. Neoadjuvant chemotherapy (NAC), defined as preoperative systemic pharmacological intervention aimed at reducing tumour burden and improving surgical resectability, is widely employed in the comprehensive management of LC. Accumulating evidence suggests that NAC can expand surgical indications, suppress tumour cell proliferation, and improve long-term survival and prognosis [3]. However, due to the limited tumour specificity of many chemotherapeutic agents, both malignant and normal cells are affected, frequently resulting in immunosuppression and severe complications such as infection, which can adversely influence treatment outcomes [4,5]. Additionally, LC patients often present with multiple comorbidities, and chemotherapy-induced myelosuppression further weakens host immune defenses. Consequently, despite its therapeutic benefits, chemotherapy markedly increases susceptibility to severe infections, which represent a significant cause of treatment interruption and failure [6]. Therefore, early identification of risk factors associated with post-chemotherapy infection and the implementation of timely, targeted preventive strategies are critical for optimizing clinical outcomes and extending patient survival [7,8].

Least Absolute Shrinkage and Selection Operator (LASSO) regression analysis was selected for its ability to effectively manage datasets containing a relatively large

number of potentially correlated predictor variables in relation to sample size. Unlike traditional stepwise regression methods, LASSO incorporates a regularization penalty that constrains regression coefficients, shrinking less informative variables toward zero and thereby achieving variable selection and model optimization. This approach reduces the risk of overfitting and improves model stability, generalizability, and interpretability [9,10]. Given the complex and multifaceted etiology of post-chemotherapy infection, as well as the need to identify the most salient predictors from numerous candidate variables, LASSO provides a robust and methodologically sound statistical framework. While previous studies have identified various infection-related risk factors in cancer patients, comprehensive and individualised prediction tools specifically designed for LC patients receiving NAC remain scarce. Furthermore, many existing predictive models rely on conventional regression methods, which may be susceptible to overfitting when handling multiple correlated predictors. Accordingly, the present study aimed to address these limitations by employing LASSO regression to develop a clinically applicable nomogram that integrates key patient-specific variables to accurately stratify the risk of post-NAC infection in LC patients.

2. Methods

2.1 Study Subjects

A total of 150 lung cancer patients who received neoadjuvant chemotherapy at Hunan Provincial People's Hospital between January 2021 and December 2024 were initially screened. Ultimately, 144 patients were included in the analysis. The sample size was determined using a consecutive sampling method during the study period. For logistic regression analysis, a commonly used rule of thumb recommends at least 10 events per variable (EPV). With 63 infection events (the dependent variable) and an initial consideration of 15 candidate predictor variables, the EPV was 4.2. Although below the ideal threshold of 10, this ratio is considered acceptable for exploratory model development, especially when combined with regularization approaches such as LASSO, which are robust in small-sample settings and can reduce overfitting when the number of events is limited [11,12]. This study was approved by the Ethics Committee of Hunan Provincial People's Hospital (Approval No. [2024]-406) and was conducted in accordance with the principles outlined in the Declaration of Helsinki. This study is a retrospective analysis of anonymized clinical data. In accordance with Article 32 of China's "Measures for Ethical Review of Life Sciences and Medical Research Involving Human Subjects" (2023), research utilising existing human data or biological samples may qualify for a waiver of informed consent provided that the study poses no harm to participants and does not involve sensitive personal information or commercial interests. Based on these criteria, the Institutional Review Board of Hunan Provincial People's Hospital waived the requirement for informed consent.

2.2 Inclusion Criteria

① Diagnosis of LC according to the World Health Organization Classification of Tumours (5th edition, Thoracic Tumours) [13] and willing to undergo chemotherapy. ② Newly diagnosed, with no prior treatment history. ③ Age $\geq$18 years. ④ Clearly defined treatment regimen, medications, and required parameters. ⑤ Availability of complete clinical data. ⑥ Completion of the planned chemotherapy course without interruption. ⑦ At the time of initial treatment, the patient or their legally authorized representative was fully informed about the chemotherapy regimen and provided written consent for the treatment.

2.3 Exclusion Criteria

① Severe trauma, systemic infection, uncontrolled severe infection, sepsis, or other severe hematological diseases. ② Presence of concurrent malignancies. ③ Renal impairment (e.g., chronic kidney disease, glomerulonephritis, kidney stones). ④ Severe autoimmune diseases (e.g., systemic lupus erythematosus or dermatomyositis). ⑤ Stress state or use of glucocorticoids or corticosteroid inhibitors within the past 3 months. ⑥ Use of immunosuppressants within the past month. ⑦ Presence of implants (e.g., pacemakers, aneurysm clips, artificial heart valves, metal fragments). ⑧ History of alcoholism, drug abuse, or substance dependence within the past 2 years.

Based on these criteria, six patients were excluded from the study: severe trauma or systemic infection (n = 1), coexisting malignancy (n = 1), renal impairment (n = 1), severe autoimmune disease (n = 1), stress or glucocorticoid use (n = 1), and immunosuppressant use (n = 1).

2.4 Data Collection

Standardized data were collected by trained researchers using predefined case report forms, which included demographic characteristics, clinical variables (such as symptoms, comorbidities, and smoking history), and laboratory indices. All data were obtained from medical records, and laboratory parameters were measured prior to the first cycle of neoadjuvant chemotherapy.

(1) Demographics and clinical characteristics: variables collected included gender, age, body mass index (BMI), Nutritional Risk Screening 2002 (NRS-2002) score, Eastern Cooperative Oncology Group Performance Status (ECOG-PS) score, Controlling Nutritional Status (CONUT) score, presence of respiratory failure, tumour stage, tumour location, pathological type, chemotherapy regimen, presence of distant metastasis, recent invasive procedures or catheterization (within 1 month), comorbidities (diabetes mellitus, chronic obstructive pulmonary disease, hypertension, hyperlipidemia), and smoking history. ① NRS-2002 [14]: This total includes impaired nutritional

status (score 1–3, higher scores indicating poorer status), disease severity (score 1–3), and an age adjustment (1 point for $\geq$70 years), yielding a total score ranging from 0 to 7. A score $\geq$3 indicates nutritional risk.

② ECOG-PS [15]: Scores ranging from 0 to 5, and classified as 0 (fully active), 1 (restricted strenuous activity), 2 (ambulatory, capable of self-care >50% waking hours), 3 (limited self-care, confined to bed/chair >50% waking hours), 4 (completely disabled), and 5 (death).

③ CONUT score [16]: Calculated using serum albumin (Alb), total cholesterol (TC), and lymphocyte count (LY), with total scores ranging from 0 to 12 and classified as normal (0–1), mild (2–4), moderate (5–8), severe (9–12).

All scale assessments (NRS-2002, ECOG-PS, CONUT) were independently performed by two trained oncology nurses. The results were subsequently reviewed and confirmed by the attending physician to ensure the accuracy of the assessment and inter-rater consistency.

(2) Laboratory parameters: laboratory data included T-cell subsets ($CD4^+$, $CD8^+$, and $CD4^+/CD8^+$ ratio), T-cell percentage ($CD3^+$), albumin (Alb), hemoglobin (Hb), neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), white blood cell count (WBC), neutrophil count (NEUT), lymphocyte count (LY), platelet count (PLT), total cholesterol (TC), red blood cell count (RBC), D-dimer, C-reactive protein (CRP), aspartate aminotransferase (AST), alanine aminotransferase (ALT), total bilirubin (TBIL), glucose (Glu), blood urea nitrogen (BUN), serum creatinine (SCr), and estimated glomerular filtration rate (eGFR).

2.5 Grouping

Patients were categorized into two groups based on the occurrence of infection during neoadjuvant chemotherapy (NAC) or within 30 days after the completion of NAC: an infection group (n = 63) and a non-infection group (n = 81). Infection was confirmed according to clinical signs and symptoms supported by radiological evidence (e.g., new infiltrates, nodules, or cavities on chest computed tomography [CT]) and/or microbiological documentation (e.g., positive sputum or blood cultures). Patients in the non-infection group completed the NAC course without any documented infectious complications.

2.6 Statistical Analysis

Statistical analyses were performed using SPSS software (version 26.0; IBM Corp., Armonk, NY, USA), R software (version 4.2.2; R Foundation for Statistical Computing, Vienna, Austria), and the Fengrui Statistics Platform (version 2.0; Fengrui Data Technology Co., Ltd., Beijing, China).

The normality of continuous variables was assessed using the Kolmogorov–Smirnov test. Normally distributed continuous variables were expressed as mean $\pm$ standard deviation ($\bar{x} \pm$ SD), and between-group comparisons were conducted using the independent-samples t-test. Non-normally distributed continuous variables were presented as median (P25, P75), and compared using the Wilcoxon rank-sum test. Categorical data are presented as [n (%)] and compared using Pearson's χ^2 test or the likelihood ratio χ^2 test, as appropriate. A two-sided p-value < 0.05 was considered statistically significant.

All pre-specified candidate predictor variables were entered into the initial model without prior univariate screening, and no variables contained missing data. For LASSO regression, continuous variables were standardized to z-scores and categorical variables were encoded as dummy variables before analysis. The optimal penalty parameter (λ) was determined via 10-fold cross-validation using the lambda.1se criterion, which selects the largest λ value within one standard error of the minimum binomial deviance, yielding a more parsimonious model. Variables with non-zero coefficients at lambda.1se were retained and subsequently entered into a multivariable logistic regression model to construct the nomogram. Nomogram construction and model validation were performed using R statistical software (Version 4.2.2; R Foundation for Statistical Computing, Vienna, Austria) and Fengrui Statistics Platform (Version 2.0; Fengrui Data Technology Co., Ltd., Beijing, China). Internal validation was performed using bootstrap resampling with 1000 iterations. Discriminative ability was evaluated by the area under the curve (AUC), with sensitivity and specificity calculated at the optimal cutoff. Model calibration was evaluated using calibration curves and the Hosmer-Lemeshow goodness-of-fit test. Clinical utility of the nomogram was evaluated using decision curve analysis (DCA) to quantify the net benefit of using the model for clinical decision-making across a range of threshold probabilities. Multicollinearity among predictors in the final multivariate model was assessed using the variance inflation factor (VIF), with VIF values <5 considered indicative of acceptable collinearity.

3. Results

3.1 Comparison of Baseline Characteristics

Compared with the non-infection group, the infection group exhibited significantly higher proportions of patients with NRS-2002 scores $\geq$3, ECOG-PS scores $\geq$2, higher CONUT scores, respiratory failure, recent invasive procedures and catheterization ($p < 0.05$). No statistically significant differences were observed between the infection group and the non-infection group in terms of gender distribution, age, BMI, tumour stage, tumour location, pathological type, chemotherapy regimen, presence of distant metastasis, presence of comorbidities (chronic obstructive pulmonary disease, hypertension, hyperlipidemia, diabetes), or smoking history (all $p > 0.05$). Detailed baseline characteristics are presented in Table 1.

Table 1. Comparison of baseline characteristics between the infection and non-infection groups.

Characteristic	Infection group (n = 63)	Non-infection group (n = 81)	$\chi^2/t/Z$ value	p-value
Sex [n (%)]			0.592	0.442
Male	56 (88.89)	75 (92.59)		
Female	7 (11.11)	6 (7.41)		
Age ($\bar{x} \pm$ SD, years)	62.16 ± 8.08	61.85 ± 7.99	–0.228	0.820
BMI (kg/m^2)	22.43 ± 3.06	22.98 ± 3.23	1.050	0.296
NRS-2002 score [n (%)]			10.465	0.001
<3 score	31 (49.21)	61 (75.31)		
≥3 score	32 (50.79)	20 (24.69)		
ECOG-PS [n (%)]			14.560	<0.001
<2	47 (74.60)	78 (96.30)		
≥2	16 (25.40)	3 (3.70)		
CONUT score	3.00 [1.00, 4.00]	2.00 [1.00, 3.00]	–3.835	<0.001
Respiratory failure [n (%)]			4.109	0.043
Yes	20 (31.75)	14 (17.28)		
No	43 (68.25)	67 (82.72)		
Tumour stage [n (%)]			6.287	0.098
I	2 (3.17)	1 (1.23)		
II	1 (1.59)	5 (6.17)		
III	12 (19.05)	26 (32.10)		
IV	48 (76.19)	49 (60.49)		
Tumour location [n (%)]			3.552	0.169
Left	20 (31.75)	37 (45.68)		
Right	37 (58.73)	35 (43.21)		
Bilateral	6 (9.52)	9 (11.11)		
Pathology [n (%)]			0.381	0.944
Small cell carcinoma	14 (22.22)	18 (22.22)		
Squamous	29 (46.03)	34 (41.98)		
Adenocarcinoma	19 (30.16)	27 (33.33)		
Other	1 (1.59)	2 (2.47)		
Chemotherapy regimen [n (%)]			0.047	0.828
Monotherapy	23 (36.51)	31 (38.27)		
Combination	40 (63.49)	50 (61.73)		
Distant metastasis [n (%)]			3.175	0.075
Yes	47 (74.60)	49 (60.49)		
No	16 (25.40)	32 (39.51)		
Recent invasive interventions/catheterization [n (%)]			7.822	0.005
Yes	34 (53.97)	25 (30.86)		
No	29 (46.03)	56 (69.14)		
Comorbidities [n (%)]				
Diabetes	13 (20.63)	11 (13.58)	1.270	0.260
COPD	17 (26.98)	18 (22.22)	0.437	0.509
Hypertension	23 (36.51)	31 (38.27)	0.047	0.828
Hyperlipidemia	9 (14.29)	11 (13.58)	0.015	0.903
Smoking history [n (%)]			0.762	0.383
Yes	45 (71.43)	63 (77.78)		
No	18 (28.57)	18 (22.22)		

Note: SD, standard deviation; BMI, body mass index; CONUT, Controlling Nutritional Status; NRS-2002, Nutritional Risk Screening 2002; ECOG-PS, Eastern Cooperative Oncology Group Performance Status; COPD, chronic obstructive pulmonary disease.

3.2 Comparison of Laboratory Parameters

Patients in the infection group showed significantly lower serum levels of $CD4^+$, $CD8^+$, the $CD4^+/CD8^+$ ratio, LY count, $CD3^+$ percentage, BUN, and SCr compared with those in the non-infection group ($p < 0.05$). In contrast, significantly higher levels of NLR, PLR, and CRP were observed in the infection group ($p < 0.05$). No statistically significant differences were found between the two

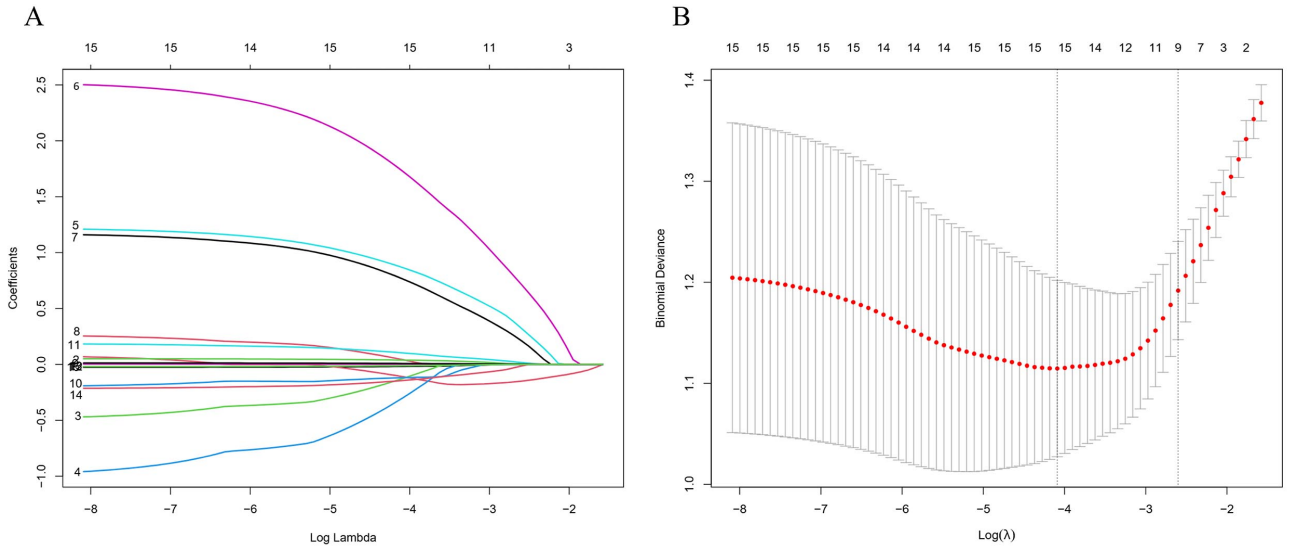

Fig. 1. LASSO regression analysis. (A) Regularization paths. (B) Ten-fold cross-validation for model selection. LASSO, Least Absolute Shrinkage and Selection Operator.

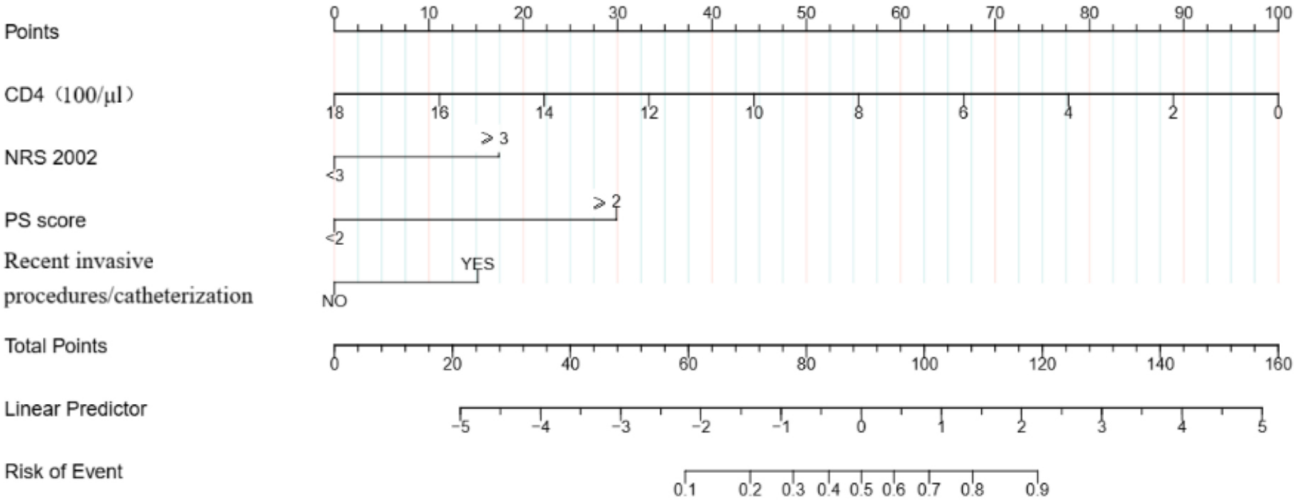

Fig. 2. Nomogram for predicting post-chemotherapy infection risk in lung cancer patients.

groups in Alb, Hb, WBC, NEUT, PLT, TC, RBC, D-dimer, AST, ALT, TBIL, Glu, or eGFR (all $p > 0.05$) (Table 2).

3.3 LASSO-Logistic Regression Analysis of Risk Factors

Post-NAC infection status was defined as the dependent variable (Y). Fifteen candidate predictors identified in Tables 1,2 were included as independent variables (X) in the LASSO regression model to select variables with non-zero coefficients. Fig. 1A illustrates the coefficient trajectories as the regularization parameter (λ) increased, and Fig. 1B presents the results of 10-fold cross-validation. The left dashed vertical line (lambda.min, $\lambda = 0.017$) indicates the value yielding the minimum binomial deviance and retained 15 variables. The right dashed line (lambda.1se, $\lambda = 0.074$) represents the value within one standard error of the minimum binomial deviance and resulted in a more parsimonious model with 9 variables. The lambda.1se ($\lambda = 0.074$) was selected for the final model, which retained 9 predictors with non-zero coefficients: NRS-2002 score, ECOG-PS, $CD4^+$ count, recent invasive procedures/catheterization, PLR, NLR, CRP, BUN, and CONUT score.

VIF analysis of the final multivariate model demonstrated no significant multicollinearity, with all variables showing VIF values <5. Subsequent multivariate logistic regression analysis identified ECOG-PS ≥ 2, recent invasive procedures/catheterization, and NRS-2002 score ≥ 3 as independent risk factors for post-NAC infection, while higher $CD4^+$ count was identified as an independent protective factor ($p < 0.05$). Detailed results are presented in Table 3.

Table 2. Comparison of laboratory parameters between infection and non-infection groups.

Parameter	Infection group (n = 63)	Non-infection group (n = 81)	t/Z value	p-value
Alb (g/L)	38.19 [34.39, 40.39]	38.28 [36.19, 40.79]	–0.461	0.643
Hb (g/L)	118.10 ± 17.34	120.15 ± 18.13	0.687	0.493
NLR	3.81 [2.81, 5.97]	2.67 [1.77, 3.61]	–4.684	<0.001
PLR	201.41 [149.67, 289.42]	133.18 [104.37, 205.04]	–4.680	<0.001
WBC ($\times 10^9$/L)	6.09 [5.20, 7.47]	6.61 [5.46, 7.90]	–0.745	0.455
NEUT ($\times 10^9$/L)	4.37 [3.24, 5.44]	4.18 [3.08, 5.21]	–1.112	0.265
LY ($\times 10^9$/L)	1.19 [0.77, 1.55]	1.52 [1.24, 2.02]	–4.176	<0.001
$CD3^+$ (%)	69.22 [58.93, 76.94]	74.29 [66.36, 80.20]	–2.267	0.023
$CD4^+$ (100 cells/µL)	3.71 [2.51, 5.30]	6.06 [4.98, 7.86]	–5.382	<0.001
$CD8^+$ (100 cells/µL)	2.99 [1.92, 3.84]	3.55 [2.89, 4.79]	–3.041	0.002
$CD4^+/CD8^+$	1.28 [1.04, 2.05]	1.76 [1.30, 2.40]	–2.588	0.010
PLT ($\times 10^9$/L)	241.00 [169.50, 309.00]	218.00 [169.00, 281.00]	–1.222	0.221
TC (mmol/L)	4.41 [3.74, 5.13]	4.58 [3.93, 5.21]	–0.983	0.325
RBC ($\times 10^{12}$/L)	3.93 ± 0.67	3.95 ± 0.73	–0.258	0.797
D-dimer (mg/L)	0.94 [0.59, 1.80]	0.73 [0.43, 1.36]	–1.492	0.135
CRP (mg/L)	13.00 [5.56, 32.40]	3.34 [3.14, 11.20]	–4.299	<0.001
ALT (U/L)	18.70 [13.80, 30.15]	20.60 [15.90, 30.20]	–0.840	0.400
AST (U/L)	23.30 [18.30, 29.55]	20.90 [17.40, 25.50]	–1.182	0.236
TBIL (µmol/L)	8.69 [6.61, 10.55]	8.18 [5.81, 10.98]	–0.459	0.645
Glu (mmol/L)	5.07 [4.75, 6.35]	5.18 [4.72, 6.37]	–0.097	0.921
BUN (mmol/L)	5.08 ± 1.70	5.76 ± 1.50	2.568	0.011
SCr (µmol/L)	63.00 [51.50, 72.00]	65.00 [58.09, 77.00]	–2.028	0.042
eGFR (mL/min)	102.86 [93.03, 107.72]	99.20 [89.48, 105.72]	–1.412	0.157

Note: Alb, albumin; ALT, alanine aminotransferase; AST, aspartate aminotransferase; BUN, blood urea nitrogen; CD, cluster of differentiation; CRP, C-reactive protein; eGFR, estimated glomerular filtration rate; Glu, glucose; Hb, hemoglobin; LY, lymphocyte count; NEUT, neutrophil count; NLR, neutrophil-to-lymphocyte ratio; PLR, platelet-to-lymphocyte ratio; PLT, platelet count; RBC, red blood cell count; SCr, serum creatinine; TBIL, total bilirubin; TC, total cholesterol; WBC, white blood cell count.

3.4 Construction of the Risk Prediction Nomogram

The risk prediction nomogram (Fig. 2) visually represents the relative contributions of the key predictors identified in the multivariate analysis, including ECOG-PS $\geq$2, recent invasive procedures/catheterization, and NRS-2002 score $\geq$3 as risk factors, and a high $CD4^+$ count as a protective factor. Points assigned to each predictor level are summed to generate a total score, which corresponds to an individualised predicted probability of post-chemotherapy infection.

3.5 Model Validation: Discrimination, Calibration, and Clinical Utility

Internal validation using bootstrap resampling (1000 iterations) demonstrated a good discriminative performance, with an AUC of 0.831 (95% confidence interval [CI]: 0.761–0.901; $p < 0.001$) (Fig. 3). At the optimal cut-off value, the model achieved a sensitivity of 76.2% and a specificity of 79.0%. Model calibration was assessed using a calibration curve (Fig. 4), which showed good agreement between predicted and observed infection probabilities. The Hosmer–Lemeshow test confirmed good calibration (χ^2 = 9.215, p = 0.324). Decision curve analysis (DCA) demonstrated that the nomogram yielded a positive net benefit across a wide range of clinically relevant threshold probabilities. Within the overall assessment range of 0.01–0.99, the model exhibited a significant net clinical benefit particularly in the intermediate threshold range of approximately 0.25–0.90, outperforming both the "treat all" and "treat none" strategies and providing superior guidance for clinical decision-making (Fig. 5).

4. Discussion

Infection is a common complication in lung cancer (LC) patients following chemotherapy. John et al. [17] demonstrated that chemotherapy-induced alterations in the tumour microenvironment can suppress the $CD4^+/CD8^+$ T-cell ratio and natural killer (NK) cell activity, thereby increasing susceptibility to opportunistic infections. This immune dysregulation, often accompanied by aberrant cytokine secretion (e.g., Interferon gamma (IFN-γ), Interleukin-6 (IL-6)), further enhances susceptibility to infection. Consequently, these pathological changes increase the incidence of pulmonary infections and other complications, significantly compromising patient prognosis. In the present study, the identification of 63 infec-

Table 3. Multivariate logistic regression analysis of risk factors for post-chemotherapy infection.

Variable	β	SE	*Wald* χ^2	*p*-value	OR	95% CI
(Intercept)	0.496	1.246	0.158	0.690	1.643	0.143–18.886
$CD4^+$	–0.268	0.108	6.155	0.013	0.765	0.619–0.945
NLR	0.144	0.144	1.002	0.317	1.155	0.871–1.531
PLR	0.002	0.003	0.501	0.479	1.002	0.996–1.009
CRP	0.011	0.014	0.632	0.427	1.011	0.983–1.040
NRS 2002 $\geq$3	1.326	0.486	7.431	0.006	3.765	1.452–9.768
ECOG-PS $\geq$2	2.291	0.829	7.640	0.006	9.883	1.947–50.152
Recent invasive procedures/catheterization	0.946	0.452	4.372	0.037	2.575	1.061–6.250
BUN	–0.300	0.159	3.553	0.059	0.741	0.542–1.012
CONUT score	0.031	0.139	0.051	0.822	1.032	0.785–1.356

CI, confidence interval; SE, standard error; OR, odds ratio.

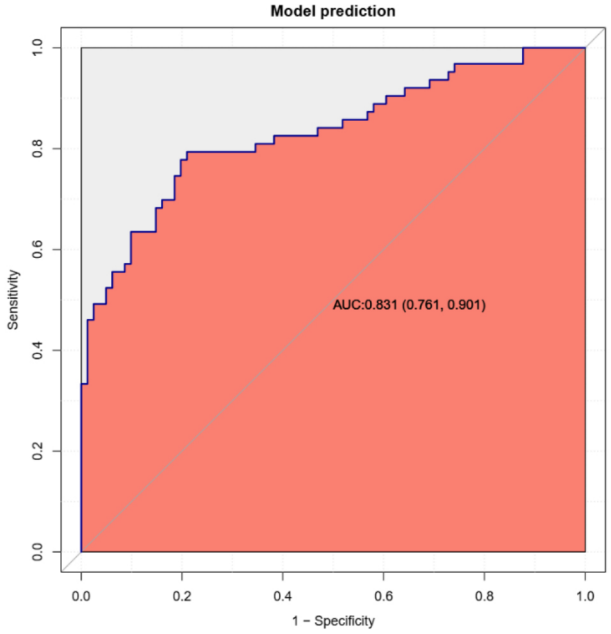

Fig. 3. ROC curve of the nomogram for predicting post-chemotherapy infection risk. ROC, receiver operating characteristic; AUC, area under the curve.

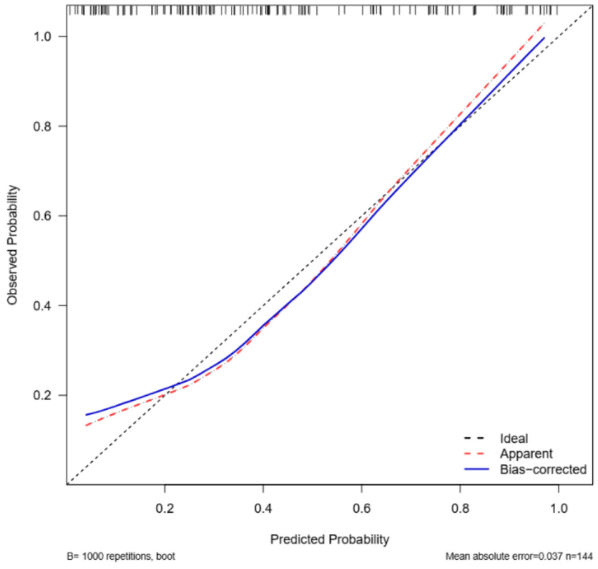

Fig. 4. Calibration curve of the LASSO-logistic prediction model.

tion events among 144 patients undergoing NAC underscores the substantial burden of post-chemotherapy infection and highlights the significance of effective preventive strategies. The complex interplay across infection, declining functional status, inflammation favouring tumour progression, and progressive immune impairment creates a vicious cycle. Therefore, early identification of high-risk patients and timely intervention are vital. Therefore, the development of an accurate and individualised prediction tool is essential for reducing infection risk and improving clinical outcomes.

LASSO regression, through the introduction of penalization to control for model complexity, enables efficient variable selection while reducing overfitting, thereby enhancing predictive accuracy and model interpretability [18,19]. Using this approach, our LASSO-logistic regression analysis identified ECOG-PS $\geq$2, recent invasive interventions/catheterization, and NRS-2002 score $\geq$3 as significant risk factors for post-NAC infection, while higher $CD4^+$ count emerged as a protective factor.

ECOG-PS $\geq$2: Performance status (PS) is a widely used measure of patients' functional capacity, and poor PS has been associated with worse survival outcomes in lung cancer patients [20], performance status (PS) assesses patients' fitness for treatment and predicts their survival. Kudo et al. [21] investigated the impact of PS on 30-day mortality in older patients with suspected infection and confirmed a significant association between higher ECOG-PS scores and pulmonary infection-related outcomes. These findings suggest that incorporating ECOG-PS monitoring into routine clinical practice may enhance risk stratification for infection and mortality in this patient population. Patients with poor PS often experience prolonged hospital stays, which increase exposure to nosocomial pathogens and elevate infection risk [22].

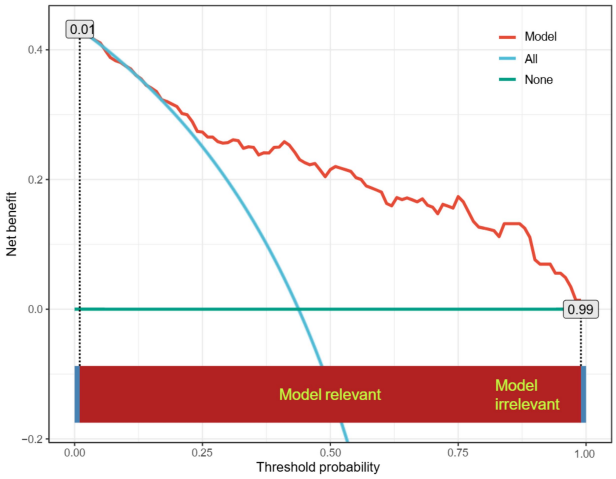

Fig. 5. Decision curve analysis of the nomogram. The y-axis represents the net benefit, reflecting the balance between true-positive and false-positive predictions. Higher net benefit values across a clinically relevant range of threshold probabilities indicate better clinical utility.

Recent invasive procedures and device placement: Invasive procedures and indwelling medical devices involve the disruption of the skin or mucosal barrier to establish access for diagnostic or therapeutic interventions, such as central venous catheterization, urinary catheterization, and endotracheal intubation. These procedures directly compromise the natural defense barriers of the body, creating potential portals of entry for pathogens. Additionally, indwelling catheters are prone to biofilm formation on their surfaces, which facilitates microbial colonization and enables evasion of host immune responses [23]. A multicenter retrospective cohort study by Wan et al. [24] reported that peripherally inserted central catheters (PICCs) significantly increase the risk of catheter-related bloodstream infections, with infection rates rising markedly after prolonged catheter dwell times exceeding 7 days. Consistent with these findings, the present study confirms that recent invasive procedures and device placement are associated with an elevated risk of post-chemotherapy infection. Furthermore, Shelke et al. [25] identified several invasive procedures, including Foley catheterization (OR 4.13), nasogastric intubation (OR 2.74), and central venous catheterization (OR 12.51), as significant risk factors for exposure to hospital-acquired multidrug-resistant (MDR) and extensively drug-resistant (XDR) organisms. These interventions prolong hospitalization, disrupt epithelial barriers, and predispose patients to severe co-infections, such as pneumonia (OR 2.10), collectively contributing to increased 30-day all-cause mortality in cancer patients.

NRS-2002 score $\geq$3: An NRS-2002 score $\geq$3 indicates a clinically significant nutritional risk characterized by weight loss, reduced dietary intake, and increased metabolic stress. These conditions impair the synthesis of proteins crucial for maintaining mucosal barrier integrity and lymphocyte function, thereby creating an "infection-malnutrition" cycle [26]. In a study investigating the nutritional status and postoperative outcomes in gastric cancer patients, Sun et al. [27] reported that individuals with an NRS score $\geq$3 who received parenteral nutrition supplementation experienced higher infection rates compared with those with scores $<$3. Furthermore, patients who developed pulmonary infections had significantly prolonged hospital stays. These findings suggest that the NRS-2002 score can effectively identify malnourished patients at high risk of infection, enabling early nutritional intervention to improve clinical prognosis. Consistently, Liu et al. [28] reported that patients with nutritional risk exhibit impaired immune function, including suppressed T-cell function, compromised mucosal barrier integrity, and reduced secretion of antimicrobial peptides. Collectively, these alterations increase susceptibility to drug-resistant bacterial infections and diminish the efficacy of anti-infective therapies.

Elevated $CD4^+$ Levels: $CD4^+$ T cells are central to cell-mediated immunity and serve as key regulators of adaptive immune responses. As a core component of host defense, $CD4^+$ T lymphocytes coordinate anti-infective immune responses through multiple mechanisms [29]: (1) assisting B-cell antibody production, whereby Th2 subsets promote the generation of pathogen-specific antibodies; (2) activating macrophages, with Th1 subsets secreting cytokines such as IFN-γ to enhance phagocytic and bactericidal activity; and (3) recruiting and activating other immune cells, including neutrophils and natural killer (NK) cells, via cytokine-mediated signaling to establish an integrated immune defense network. Therefore, higher $CD4^+$ levels reflect more robust immune surveillance and response capacity, facilitating effective pathogen clearance and thereby reducing the risk of post-chemotherapy infection.

Although univariate analysis showed significantly lower $CD8^+$ T-cell counts in the infection group, this variable was not retained in the final multivariate model. This may be explained by the strong correlation between $CD8^+$ and $CD4^+$ T-cell counts, with $CD4^+$ providing a more comprehensive indicator of overall immune competence in the predictive model. In other words, the $CD4^+$ level may serve as a more integrated indicator of overall immune competence, with declines reflecting much of the immunosuppressive information captured by reductions in $CD8^+$ T-cells. Consequently, in the multivariate analysis, the predictive effect of $CD4^+$ cells outweighed the independent contribution of $CD8^+$ cells. Wang et al. [30] reported that in patients with advanced non-small cell lung cancer (NSCLC) treated with atezolizumab, baseline $CD4^+$ T lymphocyte counts were significantly lower than those in healthy controls. This reduction, alongside an elevated $CD8^+$ T-cell count, indicated a suppressed cellular immune state. Therefore, monitoring $CD4^+$ levels and employ-

ing immunomodulatory strategies to preserve T-cells represent critical preventive strategies. In our predictive model, ECOG-PS $\geq$2 emerged as the strongest predictor of post-chemotherapy infection, followed by recent invasive procedures/catheterization, nutritional risk (NRS-2002 $\geq$3), and $CD4^+$ levels. This hierarchy underscores the combined and interrelated contributions of patient functional status, iatrogenic factors, nutritional condition, and immune status in determining infection risk. The model demonstrated robust discrimination (AUC = 0.831), excellent calibration, and clinical utility as assessed by DCA, supporting its potential application as a quantitative tool for individualised risk assessment.

Furthermore, an unexpected finding was that blood urea nitrogen (BUN) and serum creatinine (SCr) levels were lower in the infection group compared with the non-infection group. This seemingly counterintuitive observation may be explained by several factors: (1) Nutrition and muscle mass: Patients in the infection group exhibited higher NRS-2002 scores, suggesting more severe malnutrition and muscle wasting, which may reduce creatinine generation; (2) Hydration status and fluid management: Infected patients often experience volume depletion due to fever and anorexia but typically receive aggressive fluid resuscitation during clinical care, potentially causing mild hemodilution and relatively lower measured BUN and SCr levels; (3) Renal hyperfiltration: Early-phase infection may trigger inflammatory responses, transiently increasing glomerular filtration rate. Notably, there was no statistically significant difference in estimated glomerular filtration rate (eGFR) between the groups, and absolute values remained within the normal range, indicating that severe renal impairment was absent in this cohort. The underlying mechanisms driving these observations warrant further investigation in future studies.

This study has several limitations that should be acknowledged. First, its single-center, retrospective design may introduce selection bias and limit the generalizability of the findings. Second, the cohort exhibited a significant gender imbalance (92% male), which reflects the local epidemiological profile of lung cancer but restricts extrapolation of the findings to female patients and underscores the need for validation in more balanced, multicenter cohorts. Third, the model lacks external validation in an independent, multicenter population, which is essential before it can be confidently applied in broader clinical practice. Fourth, we did not explicitly analyse the impact of specific chemotherapy regimens or treatment cycles. These variables should be incorporated in larger, prospective future studies. Additionally, the model has not been directly compared with established prediction tools, such as the Multinational Association for Supportive Care in Cancer (MASCC) risk index, leaving its generalizability and comparative performance undetermined. Addressing these limitations in future research will be critical to validate the generalizability and clinical applicability of the model. Additionally, a relatively high proportion of patients with stage IV disease were included. Although these cases were documented as receiving neoadjuvant chemotherapy in the medical records, some may have received induction or conversion chemotherapy with the intent of downstaging before potential surgery, reflecting real-world clinical practice at our center. This introduces potential heterogeneity and may limit the generalizability of the findings to patients with strictly defined resectable disease. Addressing these limitations in future multicenter prospective studies will be critical to validate the generalizability and clinical applicability of the model.

5. Conclusion

In conclusion, we developed a LASSO regression-based nomogram that integrates multidimensional clinical parameters to generate individualised post-chemotherapy infection risk scores in lung cancer patients. While the model demonstrates potential for internal risk assessment, its utility for guiding stratified preventive strategies requires confirmation through external validation in independent cohorts.

Key Points

- A novel nomogram was developed and internally validated, incorporating performance status, recent invasive procedures, nutritional risk, and immune status ($CD4^+$ count) to predict individualised infection risk in patients with lung cancer undergoing neoadjuvant chemotherapy.
- The model was constructed using LASSO regression, a machine learning technique that can handle multicollinearity and select the most relevant predictors from a broad set of clinical variables, thereby enhancing robustness.
- The prediction tool demonstrated strong discriminative ability (AUC = 0.831), good calibration, and promising clinical utility across a wide range of risk thresholds, as evidenced by decision curve analysis.
- Findings of this study underscore the critical role of host factors (performance status, nutrition, immunity) and iatrogenic factors (invasive procedures) in determining infection risk, in addition to the effects of cytotoxic chemotherapy.
- The model provides a practical framework for clinicians to identify high-risk patients who may benefit from targeted preventive interventions, potentially improving treatment outcomes.
- Future multicenter, prospective studies are warranted to externally validate the model and further refine it by incorporating tumour-specific and treatment-specific variables.

Availability of Data and Materials

The data required during the current study are available upon reasonable request from the corresponding authors.

Author Contributions

LJZ designed the research study. LJZ and XJC collected the data. LJZ, XJC, JLZ, JC, JCC, JML and HY analysed the data. LJZ wrote the first draft of the manuscript. All authors contributed to the important editorial changes in the manuscript. 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

This study was approved by the Ethics Committee of Hunan Provincial People's Hospital (Approval No. [2024]-406) and was conducted in accordance with the principles outlined in the Declaration of Helsinki. Hunan Provincial People's Hospital waived the patient's informed consent. In accordance with Article 32 of China's "Measures for Ethical Review of Life Sciences and Medical Research Involving Human Subjects" (2023), research utilising existing human data or biological samples may qualify for a waiver of informed consent provided that the study poses no harm to participants and does not involve sensitive personal information or commercial interests. Based on these criteria, the Institutional Review Board of Hunan Provincial People's Hospital waived the requirement for informed consent.

Acknowledgment

Not applicable.

Funding

This study is supported by Hunan Province Clinical Medical Technology Innovation Guidance Project (2020SK50904) and Innovation Platform and Talent Plan Project of Hunan Provincial Department of Science and Technology (No. 2023SK4056).

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians. 2021; 71: 209–249. https://doi.org/10.3322/caac.21660

[2] Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians. 2024; 74: 229–263. https://doi.org/10.3322/caac.21834

[3] Jeon H, Gor R, D'Aiello A, Stiles B, Illei PB, Halmos B. Advancing neoadjuvant therapies in resectable non-small cell lung cancer: implications for novel treatment strategies and biomarker discovery. Pathology and Oncology Research. 2024; 30: 1611817. https://doi.org/10.3389/pore.2024.1611817

[4] Da Silva R, Casella T. Healthcare-associated infections in patients who are immunosuppressed due to chemotherapy treatment: a narrative review. Journal of Infection in Developing Countries. 2022; 16: 1784–1795. https://doi.org/10.3855/jidc.16495

[5] Nayak L, Salunke G, Gilada T, Savarkar S, Salunke B, Biswas S, et al. Expert insights: 10 key questions on managing common infections in cancer care in India. Ecancermedicalscience. 2025; 19: 1971. https://doi.org/10.3332/ecancer.2025.1971

[6] Bertaglia V, Morelli AM, Solinas C, Aiello MM, Manunta S, Denaro N, et al. Infections in lung cancer patients undergoing immunotherapy and targeted therapy: An overview on the current scenario. Critical Reviews in Oncology/Hematology. 2023; 184: 103954. https://doi.org/10.1016/j.critrevonc.2023.103954

[7] Sun T, Liu J, Yuan H, Li X, Yan H. Construction of a risk prediction model for lung infection after chemotherapy in lung cancer patients based on the machine learning algorithm. Frontiers in Oncology. 2024; 14: 1403392. https://doi.org/10.3389/fonc.2024.1403392

[8] Qin W, You T, Guo T, Tian R, Cui X, Wang P. The Inter-Relationships Among the Risk Factors for Pulmonary Infection and the Diagnostic Utility of Inflammatory Markers in Patients with Non-Small Cell Lung Cancer. Infection and Drug Resistance. 2025; 18: 1111–1123. https://doi.org/10.2147/IDR.S491507

[9] Fuetterer C, Nalenz M, Augustin T, Pfeiffer RM. A powerful penalized multinomial logistic regression approach. Computational Statistics. 2025; 40: 4565–4587. https://doi.org/10.1007/s00180-025-01635-0

[10] Ma Z, Yang S, Yang Y, Luo J, Zhou Y, Yang H. Development and validation of prediction models for the prognosis of colon cancer with lung metastases: a population-based cohort study. Frontiers in Endocrinology. 2023; 14: 1073360. https://doi.org/10.3389/fendo.2023.1073360

[11] Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological). 1996; 58: 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

[12] Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. American Journal of Epidemiology. 2007; 165: 710–718. https://doi.org/10.1093/aje/kwk052

[13] WHO classification of tumours editorial board. Thoracic tumours. International Agency for Research on Cancer: Lyon. 2021.

[14] Kondrup J, Allison SP, Elia M, Vellas B, Plauth M, Educational and Clinical Practice Committee, European Society of Parenteral and Enteral Nutrition (ESPEN). ESPEN guidelines for nutrition screening 2002. Clinical Nutrition. 2003; 22: 415–421. https://doi.org/10.1016/s0261-5614(03)00098-0

[15] Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. American Journal of Clinical Oncology. 1982; 5: 649–655. https://doi.org/10.1097/00000421-198212000-00014

[16] Ignacio de Ulíbarri J, González-Madroño A, de Villar NGP, González P, González B, Mancha A, et al. CONUT: a tool for controlling nutritional status. First validation in a hospital population. Nutricion Hospitalaria. 2005; 20: 38–45.

[17] John L, Miah K, Benner A, Mai EK, Kriegsmann K, Hundemer M, et al. Impact of novel agent therapies on immune cell subsets and infectious complications in patients with relapsed/refractory

multiple myeloma. Frontiers in Oncology. 2023; 13: 1078725. https://doi.org/10.3389/fonc.2023.1078725

[18] Kipruto E, Sauerbrei W. Evaluating Prediction Performance: A Simulation Study Comparing Penalized and Classical Variable Selection Methods in Low-Dimensional Data. Applied Sciences. 2025; 15: 7443. https://doi.org/10.3390/app15137443

[19] Zhou DJ, Chahal R, Gotlib IH, Liu S. Comparison of Lasso and Stepwise Regression in Psychological Data. Methodology. 2024; 20: 121–143. https://doi.org/10.5964/meth.11523

[20] Cunha MT, de Souza Borges AP, Carvalho Jardim V, Fujita A, de Castro G, Jr. Predicting survival in metastatic non-small cell lung cancer patients with poor ECOG-PS: A single-arm prospective study. Cancer Medicine. 2023; 12: 5099–5109. https://doi.org/10.1002/cam4.5254

[21] Kudo M, Sasaki S, Takada T, Fujii K, Yagi Y, Yano T, et al. Predicting 30-day mortality in older patients with suspected infections by adding performance status to quick sequential organ failure assessment. Journal of General and Family Medicine. 2025; 26: 238–245. https://doi.org/10.1002/jgf2.764

[22] Meyers DE, Pasternak M, Dolter S, Grosjean HAI, Lim CA, Stukalin I, et al. Impact of Performance Status on Survival Outcomes and Health Care Utilization in Patients With Advanced NSCLC Treated With Immune Checkpoint Inhibitors. JTO Clinical and Research Reports. 2023; 4: 100482. https://doi.org/10.1016/j.jtocrr.2023.100482

[23] Bouhrour N, Nibbering PH, Bendali F. Medical Device-Associated Biofilm Infections and Multidrug-Resistant Pathogens. Pathogens. 2024; 13: 393. https://doi.org/10.3390/pathogens13050393

[24] Wan R, Gu L, Yin B, Cai S, Zhou R, Yang W. A six-year study of complications related to peripherally inserted central catheters: A multi-center retrospective cohort study in China. Perfusion. 2023; 38: 689–697. https://doi.org/10.1177/02676591221076287

[25] Shelke A, Priya P, Mishra S, Chauhan R, Murti K, Ravichandiran V, et al. Investigation of antimicrobial susceptibility patterns, risk factors and their impact on mortality in cancer patients at a tertiary care cancer hospital- A prospective study. Annals of Clinical Microbiology and Antimicrobials. 2024; 23: 59. https://doi.org/10.1186/s12941-024-00703-5

[26] Yang ZY, Yang F. Nutritional status of patients with gastrointestinal cancers and analysis of factors for postoperative infections. BMC Cancer. 2024; 24: 1389. https://doi.org/10.1186/s12885-024-13093-w

[27] Sun S, Sun W, Xie W, Zhao F, Guo X, Zhou J, et al. The effect of supplementary parenteral nutrition with different energy intakes on clinical outcomes of patients after gastric cancer surgery. BMC Surgery. 2024; 24: 424. https://doi.org/10.1186/s12893-024-02734-3

[28] Liu J, Liu J, Chen H, Zhang Q, Zhang P, Chen Z, et al. Prognostic value of combined nutritional and inflammatory markers in NSCLC patients receiving ICIs. Discover Oncology. 2025; 16: 571. https://doi.org/10.1007/s12672-025-02391-9

[29] Lee GR. Molecular Mechanisms of T Helper Cell Differentiation and Functional Specialization. Immune Network. 2023; 23: e4. https://doi.org/10.4110/in.2023.23.e4

[30] Wang Y, Zhu J, Zhou N, Wang Y, Zhang X. Changes in T lymphocyte subsets predict the efficacy of atezolizumab in advanced non-small cell lung cancer: a retrospective study. Journal of Thoracic Disease. 2023; 15: 5669–5679. https://doi.org/10.21037/jtd-23-1169