Am J Cancer Res 2026;16(7):2707-2724
www.ajcr.us /ISSN:2156-6976/ajcr0173689

Original Article

Incidence and risk factors of infections in lung cancer patients receiving immunotherapy combined with chemotherapy: a real-world retrospective study

Jing Zou[1], Guanghai Ji[2], Lei Yan[1], Jinyong Li[1], Yanxia Zhang[1], Dingfang Pi[3]

[1]Pulmonary and Critical Care Medicine, The First Affiliated Hospital of Yangtze University/The First People's Hospital of Jingzhou, Jingzhou 434000, Hubei, China; [2]Department of Radiology, The First Affiliated Hospital of Yangtze University/The First People's Hospital of Jingzhou, Jingzhou 434000, Hubei, China; [3]Department of Infectious Diseases, The First Affiliated Hospital of Yangtze University/The First People's Hospital of Jingzhou, Jingzhou 434000, Hubei, China

Received April 29, 2026; Accepted June 14, 2026; Epub July 15, 2026; Published July 30, 2026

Abstract: This single-center, real-world retrospective cohort study assessed the incidence, risk factors, and clinical risk stratification of infections in lung cancer (LC) patients receiving immune checkpoint inhibitors (ICIs) combined with chemotherapy. This study comprised a cohort of 616 consecutive LC patients who received ICI-based chemotherapy during hospitalization between January 2020 and December 2025. LASSO regression followed by multivariable logistic regression was used to identify independent risk factors. Optimal cutoffs for score construction were determined using receiver operating characteristic (ROC) curve analysis combined with the Youden index, and internal validation was performed with 1,000 bootstrap resamples. The overall infection rate was 9.09% (56/616), with pulmonary infection being the most common site (64.29%), and gram-negative bacteria being most predominant (50.00%) Among culture-positive specimens, CRKP, MRSA, and CRPA accounted for 40.00%, 33.33%, and 25.00% of their respective species. Multivariable analyses identified ECOG performance status ≥2 (OR=3.993, 95% CI: 1.520-10.926, P=0.005), C-reactive protein (CRP) (OR=1.114, 95% CI: 1.079-1.159, P<0.05) as independent risk factors. In summary, infections appear to be a major complication in LC patients receiving ICIs combined with chemotherapy, often involving drug-resistant pathogens. A scoring system based on seven routinely available clinical parameters demonstrated good discrimination and internal stability in bootstrap validation, potentially aiding in the identification of high-risk patients for preventive measures.

Keywords: Lung cancer, immune checkpoint inhibitors, infection, risk factors, prediction model

Introduction

Lung cancer (LC) is one of the most lethal malignancies worldwide. According to GLOBOCAN 2022, Bray et al. [1] stated approximately 2.5 million new cases in 2022, accounting for 12.4% of all new cancer diagnoses worldwide, and about 1.8 million deaths (18.7%). Both the incidence and mortality of lung cancer were the highest among all cancers, representing a significant public health problem. In China, Han et al. [2] found that lung cancer ranked first in both incidence and mortality in 2022. Due to limited use of early screening, most patients present with locally advanced or metastatic diseases, and the overall prognosis remains poor.

The treatment landscape of lung cancer has been profoundly transformed by immune checkpoint inhibitors (ICIs), specifically programmed death receptor 1 (PD-1) and programmed death-ligand 1 (PD-L1) inhibitors. ICIs combined with platinum-based chemotherapy significantly improve progression-free survival (PFS) and overall survival (OS) in advanced non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC), and have become the standard first-line therapy according to the national and international guidelines. In the phase 3 CASPIAN trial [3], Paz-Ares et al. reported that the addition of durvalumab to platinum-etoposide improved OS compared with chemotherapy alone (13.0 vs. 10.3 months; HR=0.73, P=0.0047) in patients with extensive-stage

https://doi.org/10.62347/OIJD3060

SCLC, establishing immunochemotherapy as the first-line standard of care. Real-world data also support the survival benefit of ICIs: pembrolizumab-treated patients with advanced NSCLC had a significantly longer median OS than patients treated with chemotherapy alone (13.0 vs. 9.2 months) [4]. However, adverse event and hospitalization rates were higher in the immunotherapy group, implying that survival gains must be accompanied by effective management of complications.

Lung cancer patients are prone to immunosuppression due to the tumor itself and often accompanied by malnutrition and declining performance status [5]. Chemotherapy-induced myelosuppression further reduces neutrophils and lymphocytes and lowers serum albumin, thus impairing host defenses. The infection risk associated with ICIs is complex [6]: most evidence indicates that infections are more likely associated with immune-related adverse events (irAEs) and the corticosteroids or other immunosuppressants used to manage irAEs, rather than to ICI per se [7]. The profile of infective risks under immunochemotherapy is therefore more complicated than with chemotherapy alone.

Infections can delay tumor therapy, prolong hospital stay, which increase medical costs. For advanced NSCLC, direct medical costs may range from tens to hundreds of thousands of US dollars, with hospitalization accounting for 5.0%-58.1% of total costs [8] and infectious complications further escalating expenses. Lower respiratory tract infections remain the leading cause of infectious mortality worldwide, with around 2.5 million deaths expected in 2023 [9]. In cancer patients receiving immunosuppressive therapy, infection-related mortality risk is markedly elevated. Therefore, effective strategies to prevent infections in this patient population are of utmost importance.

However, few studies have specifically investigated infections in lung cancer patients undergoing immunochemotherapy. The majority of available literature focuses on the identification and management of irAEs rather than infections [10]. Current reports on infection are mainly small, single-center, retrospective studies with highly heterogeneous populations, varying definitions of infection, different observation periods, and widely disparate incidence rates, limiting their comparability and generalizability [11]. Risk factor analyses, as reported to date, have largely been univariate, and multivariate analyses addressing ICI-based combination therapy are lacking. A meta-analysis by Chen et al. [12] on risk factors for pulmonary infections in NSCLC patients synthesized data from 10 studies and demonstrated that diabetes and hypoalbuminemia may be relevant predictors. However, none of the studies included patients receiving ICI-based combination treatment. Consequently, guidance for identifying high-risk patients and implementing preventive strategies are lacking. At present, no validated clinical prediction tool exists for this population.

In light of these gaps, we conducted a retrospective cohort study using real-world clinical data from LC patients who received ICIs combined with chemotherapy. Our goals were to (1) characterize the incidence, site distribution, and pathogen profile of infections, (2) identify independent risk factors using univariate/multivariable logistic regression, (3) develop a simple, clinically usable risk scoring system based on LASSO-based variable selection and ROC-derived optimal cut-offs, and (4) evaluate its stability and predictive performance through bootstrap internal validation. The resulting tool is intended to help clinician identify patients at high risk early, allowing rational allocation of preventive interventions to reduce infection rates and improve treatment tolerability and prognosis.

Materials and methods

Study design and participants

This retrospective study was conducted at the First Affiliated Hospital of Yangtze University (The First People's Hospital of Jingzhou). Using the hospital's electronic medical record system, we consecutively collected clinical data from lung cancer patients admitted between January 1, 2020 and December 31, 2025, who received ICIs combined with chemotherapy. Among the 616 patients enrolled, 56 (9.09%) developed an infection. The study was approved by the Ethics Committee of the First Affiliated Hospital of Yangtze University (The First People's Hospital of Jingzhou). Due to the retrospective design, informed consent was waived per relevant regulations, and all data

collection and use were consistent with the Declaration of Helsinki (Figure S1). The sample size was determined based on a previous report indicating that the incidence of infection in LC patients receiving immunotherapy ranges from 22% to 27%. We used a midpoint estimate of 24.5% [13]. According to a single-proportion formula for two-sided α=0.05 (Z=1.96) and margin of error E=0.05, the minimum required sample size was approximately 284. Considering a loss/dropout rate of 10%, the adjusted minimum sample size was approximately 316. Our study finally enrolled 616 patients, exceeding this requirement.

Inclusion criteria: (1) pathology- or cytology-proven lung cancer (NSCLC or SCLC); (2) aged 18 years or older; (3) receipt of at least one cycle of ICI-based combination chemotherapy at our center; (4) inpatient admission for at least 24 hours; (5) sufficiently complete clinical records for retrospective analysis. Exclusion criteria: (1) concurrent diagnosis of other malignancy; (2) active infection on admission; (3) prior organ transplantation or long-term use of immunosuppressants; (4) missing crucial data; or (5) incomplete follow-up data.

Definition of infection

Infection was defined as a new-onset clinically or microbiologically confirmed infectious event that occurred during hospitalization and treatment. The diagnosis was based on clinical manifestations (e.g., fever, cough, dyspnea), laboratory findings (e.g., elevated inflammatory markers), imaging and microbiological evidence (when available) [14]. Patients showing signs of active infection at admission or baseline were excluded. Cases presenting with irAEs (e.g., immune-related pneumonitis), without evidence of infection, were excluded.

Clinical data collection

Data were extracted from the electronic medical record system. Demographic and baseline characteristics included age, sex, body mass index (BMI), smoking history, and other general clinical information. Tumor-related variables included histological type (adenocarcinoma, squamous cell carcinoma, SCLC, and other), disease stage (per the 8th edition of the AJCC TNM staging system), PD-L1 expression level (<1%, 1%-49%, ≥50%), driver gene mutation status (e.g., *EGFR* and *ALK*), and Eastern Cooperative Oncology Group (ECOG) performance score.

Recorded comorbidities included diabetes, hypertension, chronic obstructive pulmonary disease (COPD), chronic heart disease, chronic renal insufficiency, and chronic liver disease. Treatment-related variables included ICI type (PD-1 or PD-L1 inhibitor), chemotherapy regimen (platinum-doublet or single-agent), treatment line (first-line or second-line and beyond), number of completed chemotherapy cycles, corticosteroid use, and prophylactic use of antibiotics and granulocyte colony-stimulating factor (G-CSF).

Laboratory parameters collected during treatment included white blood cell (WBC) count, neutrophil count, lymphocyte count, serum albumin, hemoglobin, and C-reactive protein (CRP). Neutropenia was defined as a neutrophil count $<1.5\times10^9$/L [15]; hypoalbuminemia as serum albumin <35 g/L [16]; and lymphocytopenia as a lymphocyte count $<1.0\times10^9$/L [16]. All laboratory data were obtained from the most recent measurement before the start of the current treatment cycle or before the onset of infection. To ensure baseline comparability, blood samples were routinely collected within 24-48 hours of hospital admission.

Measurement methods

All laboratory tests were performed in the clinical laboratory of our institution using standardized methods. Complete blood counts (including WBC, neutrophil count, lymphocyte count, and hemoglobin) were measured on a Sysmex XN-3000 automated hematology analyzer (Sysmex Corporation, Japan) with manufacturer-supplied reagent kits. Daily internal quality control was performed before each run. Serum albumin levels were measured by the bromocresol green colorimetry on a Beckman Coulter AU5800 automated biochemistry analyzer (Beckman Coulter, USA) with corresponding reagent kits. Serum CRP levels were measured using commercially available enzyme-linked immunosorbent assay (ELISA) kit (Catalog No. ml300545, Shanghai Enzyme-linked Biotechnology Co., Ltd., Shanghai, China) according to the manufacturer's instructions. All laboratory procedures were performed in accordance with the standard oper-

ating procedures provided by the manufacturer. Laboratory personnel received standardized training, and all test results met institutional and external quality-control requirements.

Outcomes

The primary outcome was the occurrence of infection during hospitalization, defined as a new infection event diagnosed according to standardized clinical and microbiological criteria, with reference to established diagnostic criteria, including the Chinese Diagnostic Criteria for Nosocomial Infections (2001). Infection site distribution, pathogen profile, and antimicrobial resistance patterns were also analyzed.

Secondary outcomes included identification and quantification of independent risk factors for infection, evaluation of the predictive performance of the clinical risk scoring system (AUC, sensitivity, specificity), assessment of model stability through 1,000-bootstrap resampling, and examination of score consistency across subgroups by histological type, ICI class, treatment line, and ECOG status.

Statistical analysis

All analyses were performed using R (version 4.5.2, R Core Team) and SPSS (version 27.0, IBM Corporation). Normality of continuous variables was assessed using the Shapiro-Wilk test (Table S1). Normally distributed continuous variables were expressed as mean ± standard deviation (SD) and compared using independent-samples t-test. Non-normally distributed variables were expressed as median (interquartile range) [P25, P75] and compared using Mann-Whitney U test. Categorical variables were presented as counts and proportions (%) and compared using chi-square test, or Fisher exact test when expected cell frequencies were <5.

Before multivariable analysis, candidate variables were assessed for multicollinearity using Spearman correlation and variance inflation factors (VIFs). Variables with $P<0.05$ in the univariate analysis were considered candidate predictors for subsequent modeling. To reduce redundancy before least absolute shrinkage and selection operator (LASSO) regression, three binary indicators - neutropenia, hypoalbuminemia, and lymphocytopenia - were excluded because they overlapped with their corresponding continuous variables (namely neutrophil count, serum albumin, and lymphocyte count, respectively).

A total of 14 variables were entered into a LASSO regression model with L1 regularization. The optimal penalty parameter λ was determined using 10-fold cross-validation, and the lambda.1se criterion was selected to obtain a more parsimonious and stable model. Variables retained at lambda.1se were entered into multivariable logistic regression, and age ≥65 years was additionally included based on clinical relevance. Multicollinearity among variables selected for modeling was reassessed using VIF, and all VIFs were below 3, indicating no substantial collinearity.

The final multivariable model included 11 candidate variables and 56 infection events, yielding an events-per-variable (EPV) ratio of approximately 5.1, which was below the conventional 10:1 threshold. Therefore, LASSO-based dimension reduction and 1,000-bootstrap internal validation were used to reduce and assess the risk of model overfitting. Results were expressed as odds ratios (ORs) with 95% confidence intervals (CIs).

For continuous variables that remained independently associated with infection in the multivariable model, ROC curve analysis was performed, and optimal cutoffs were determined using the Youden index for risk score development. The predictive performance of the scoring system was evaluated using AUC, sensitivity, specificity, and Brier score based on 1,000 bootstrap resamples. AUCs were compared using the DeLong test, while net reclassification improvement (NRI) and integrated discrimination improvement (IDI) were calculated in sensitivity analyses to assess incremental predictive value.

Subgroup analyses were performed according to histological type, ICI class, treatment line, and ECOG status, with interaction tests performed to assess consistency of the scoring system across subgroups. To minimize reverse causation, only pre-infection laboratory measurements were included for patients who developed infection. All tests were two-sided, and $P<0.05$ was considered statistically significant.

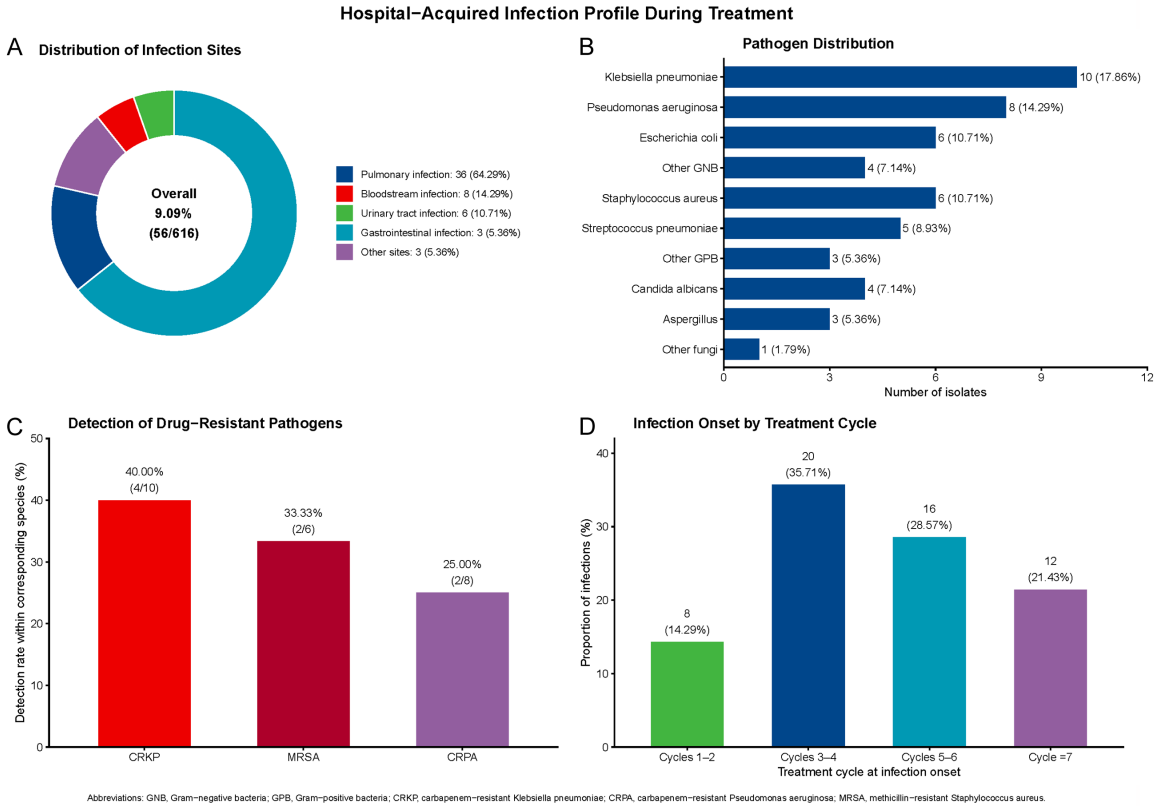

Figure 1. Distribution of infection characteristics in lung cancer patients receiving ICI-based chemotherapy. A. Distribution of infection sites; B. Pathogen profile; C. Detection rates of drug-resistant organisms among their respective species; D. Distribution of infection events by chemotherapy cycle at onset. Note: GNB, Gram-negative bacteria; GPB, Gram-positive bacteria; CRKP, carbapenem-resistant Klebsiella pneumoniae; CRPA, carbapenem-resistant Pseudomonas aeruginosa; MRSA, methicillin-resistant Staphylococcus aureus.

Results

Incidence, type, and pathogen distribution of infections

Among the 616 patients receiving ICI-based chemotherapy, 56 developed an infection, with an overall incidence of 9.09% (95% CI: 6.9-11.7%). Pulmonary infection was the most common site, affecting more than 60% of infected patients, followed by bloodstream infection, urinary tract infection, gastrointestinal infection, and other sites infections (**Figure 1A**).

Gram-negative bacteria accounted for 50.00% of pathogens, with *Klebsiella pneumoniae* (17.86%), *Pseudomonas aeruginosa* (14.29%), and *Escherichia coli* (10.71%) being the most frequent pathogens. Gram-positive species accounted for 25.00% of isolates, predominantly *Staphylococcus aureus* (10.71%) and *Streptococcus pneumoniae* (8.93%). Fungal infection was identified in 14.29% of the cases, including *Candida albicans* (7.14%) and *Aspergillus* spp. (5.36%). In 10.71% of cases, cultures were negative or did not identify a pathogen (**Figure 1B**).

Among drug-resistant organisms, carbapenem-resistant *K. pneumoniae* (CRKP) accounted for 40.00% (4/10), methicillin-resistant *S. aureus* (MRSA) for 33.33% (2/6), and carbapenem-resistant *P. aeruginosa* (CRPA) for 25.00% (2/8) (**Figure 1C**). Infections occurred most common during chemotherapy cycles 3-4 (35.71%), followed by cycles 5-6 (28.57%), with a declining trend from cycle 7 onwards (21.43%). The lowest incidence was observed in cycles 1-2 (14.29%) (**Figures 1D**, S2).

Baseline characteristics

Comparison of baseline characteristics between infection and non-infection groups showed significant differences in age ($P< 0.001$), proportion of patients aged ≥65 years

(P=0.004), BMI (P<0.001), ECOG score (P< 0.001), and tumor stage (P=0.020). Comorbidities such as diabetes (P=0.034) and COPD (P=0.009) were more prevalent in the infected group, while overall comorbidity burden didn't differ significantly (P=0.102). Significant differences were observed in laboratory parameters including WBC and albumin (both P<0.001). Treatment-related factors showing significant differences included number of completed chemotherapy cycles (P=0.024), neutropenia (P<0.001), hypoalbuminemia (P<0.001), lymphocytopenia (P<0.001), and corticosteroid use (P<0.001). No significant differences were observed between groups in terms of sex (P=0.409), smoking history (P=0.201), histological type (P=0.712), PD-L1 expression (P=0.826), driver mutation status (P=0.434), hypertension (P=0.376), chronic heart disease (P=0.316), chronic renal insufficiency (P= 0.216), chronic liver disease (P=0.583), ICI type (P=0.709), chemotherapy regimen (P= 0.949), treatment line (P=0.151), prior radiotherapy (P=0.051), prior surgery (P=0.701), prophylactic G-CSF use (P=0.132), or prophylactic antifungal use (P=0.316) (all P>0.05). Details are shown in **Table 1**.

Multicollinearity assessment

Spearman correlation analysis of 17 candidate variables showed moderate correlations between lymphocyte count and lymphocytopenia (r=-0.780, P<0.001), WBC and neutrophil count (r=0.773, P<0.001), and serum albumin and hypoalbuminemia (r=-0.758, P<0.001). All other pairwise correlations fell below 0.6, indicating no strong linear relationships (**Figure 2A**). VIF values were all below 5, confirming the absence of severe multicollinearity; all variables were retained for subsequent regression analysis (**Figure 2B**).

LASSO variable selection

LASSO regression was applied to 17 variables that reached statistical significance in univariate analysis. To avoid redundancy, three binary indicators - neutropenia, hypoalbuminemia, and lymphocytopenia - were removed because they represented the same clinical dimensions as their corresponding continuous variables (neutrophil count, albumin, and lymphocyte count). Fourteen variables were retained for LASSO-based selection. Ten-fold cross-validation identified the minimum λ (lambda.min=0.0029), at which white only WBC count was excluded, leaving 13 variables. Using the more parsimonious lambda.1se=0.0116, age ≥65 years, stage IV, and diabetes were additionally excluded, leaving 10 variables for final selection: BMI, ECOG score ≥2, COPD, neutrophil count, lymphocyte count, albumin, hemoglobin, CRP, completed chemotherapy cycles, and corticosteroid use (**Figure 3A**, **3B**). The lambda.1se solution was selected to prioritize model parsimony and predictive stability.

Multivariable logistic regression

The 10 variables retained by LASSO at lambda.1se, together with age ≥65 years, which was additionally included based on clinical judgment, yielded 11 candidate predictors for multivariable logistic regression. VIF reassessment among these variables showed all values below 3, confirming no substantial multicollinearity. Multivariable analysis identified ECOG score ≥2 (OR=4.180, 95% CI: 1.564-11.736, P=0.005), CRP (OR=1.120, 95% CI: 1.083-1.166, P<0.001), and number of completed chemotherapy cycles (OR=1.290, 95% CI: 1.017-1.669, P=0.042) as independent risk factors for infection. In contrast, neutrophil count (OR=0.629, 95% CI: 0.471-0.815, P<0.001), lymphocyte count (OR=0.174, 95% CI: 0.060-0.450, P<0.001), albumin (OR= 0.820, 95% CI: 0.734-0.907, P<0.001), and hemoglobin (OR=0.942, 95% CI: 0.909-0.973, P<0.001) were independent protective factors. Age ≥65 years (P=0.770), BMI (P=0.157), COPD (P=0.081), and corticosteroid use (P=0.071) did not reach statistical significance (**Figure 4**).

ROC analysis of continuous independent variables

ROC curves were plotted for the six continuous variables independently associated with infection in the multivariable model. Infection status served as the binary outcome and optimal cutoffs were determined using the Youden index. CRP had the highest discriminative ability, with an area under the curve (AUC) of 0.841 (95% CI: 0.775-0.906) and an optimal cutoff of 33.025 mg/L (sensitivity 67.9%, specificity 91.4%) (**Figure 5E**). Albumin (AUC=0.771, 95% CI: 0.701-0.841) and lymphocyte count (AUC= 0.769, 95% CI: 0.711-0.828) showed similar predictive performance. The optimal cutoffs

Table 1. Comparison of baseline characteristics between infection and non-infection groups in lung cancer patients receiving immunotherapy combined with chemotherapy

Variable	Category	Total (N=616)	Non-infection group (n=560)	Infection group (n=56)	Statistic	*P* value
Age (years)	Median [P25, P75]	64.00 [58.00, 70.00]	64.00 [58.00, 69.00]	69.00 [60.75, 77.00]	4.091	<0.001
Age ≥65 years	Yes	304 (49.35%)	266 (47.50%)	38 (67.86%)	8.440	0.004
	No	312 (50.65%)	294 (52.50%)	18 (32.14%)		
Sex	Male	515 (83.60%)	466 (83.21%)	49 (87.50%)	0.682	0.409
	Female	101 (16.40%)	94 (16.79%)	7 (12.50%)		
BMI (kg/m^2)	Mean ± SD	22.21±3.56	22.37±3.57	20.60±3.08	3.572	<0.001
Smoking history	Yes	335 (54.38%)	300 (53.57%)	35 (62.50%)	1.636	0.201
	No	281 (45.62%)	260 (46.43%)	21 (37.50%)		
ECOG score	0-1	448 (72.73%)	420 (75.00%)	28 (50.00%)	16.042	<0.001
	≥2	168 (27.27%)	140 (25.00%)	28 (50.00%)		
Tumor stage	Stage III	207 (33.60%)	196 (35.00%)	11 (19.64%)	5.381	0.020
	Stage IV	409 (66.40%)	364 (65.00%)	45 (80.36%)		
Histological type	Adenocarcinoma	305 (49.51%)	280 (50.00%)	25 (44.64%)	1.373	0.712
	Squamous cell carcinoma	216 (35.06%)	196 (35.00%)	20 (35.71%)		
	SCLC	75 (12.18%)	67 (11.96%)	8 (14.29%)		
	Other	20 (3.25%)	17 (3.04%)	3 (5.36%)		
PD-L1 expression	<1%	214 (34.74%)	196 (35.00%)	18 (32.14%)	0.383	0.826
	1%-49%	246 (39.94%)	224 (40.00%)	22 (39.29%)		
	≥50%	156 (25.32%)	140 (25.00%)	16 (28.57%)		
Driver mutation status	Positive (EGFR/ALK, etc.)	182 (29.55%)	168 (30.00%)	14 (25.00%)	0.611	0.434
	Negative	434 (70.45%)	392 (70.00%)	42 (75.00%)		
Comorbidities (any)	Yes	366 (59.42%)	327 (58.39%)	39 (69.64%)	2.672	0.102
	No	250 (40.58%)	233 (41.61%)	17 (30.36%)		
Diabetes	Yes	130 (21.10%)	112 (20.00%)	18 (32.14%)	4.508	0.034
	No	486 (78.90%)	448 (80.00%)	38 (67.86%)		
Hypertension	Yes	188 (30.52%)	168 (30.00%)	20 (35.71%)	0.784	0.376
	No	428 (69.48%)	392 (70.00%)	36 (64.29%)		
COPD	Yes	100 (16.23%)	84 (15.00%)	16 (28.57%)	6.895	0.009
	No	516 (83.77%)	476 (85.00%)	40 (71.43%)		
Chronic heart disease	Yes	64 (10.39%)	56 (10.00%)	8 (14.29%)	1.004	0.316
	No	552 (89.61%)	504 (90.00%)	48 (85.71%)		
Chronic renal insufficiency	Yes	46 (7.47%)	39 (6.96%)	7 (12.50%)	1.528	0.216
	No	570 (92.53%)	521 (93.04%)	49 (87.50%)		
Chronic liver disease	Yes	39 (6.33%)	34 (6.07%)	5 (8.93%)	0.302	0.583
	No	577 (93.67%)	526 (93.93%)	51 (91.07%)		
WBC count (×10^9/L)	Mean ± SD	5.76±2.30	5.86±2.29	4.75±2.14	3.456	<0.001
Neutrophil count (×10^9/L)	Median [P25, P75]	3.92 [2.53, 5.30]	4.08 [2.78, 5.41]	2.46 [1.37, 4.21]	4.718	<0.001
Lymphocyte count (×10^9/L)	Mean ± SD	1.32±0.57	1.37±0.56	0.84±0.44	6.872	<0.001
Albumin (g/L)	Mean ± SD	38.03±5.01	38.50±4.73	33.31±5.40	7.727	<0.001
Hemoglobin (g/L)	Mean ± SD	111.08±17.99	112.51±17.53	96.80±16.32	6.430	<0.001
CRP (mg/L)	Median [P25, P75]	15.34 [6.90, 26.36]	14.34 [6.26, 23.25]	47.06 [22.96, 65.27]	8.412	<0.001
Completed chemotherapy cycles	Median [P25, P75]	4.00 [3.00, 5.00]	4.00 [3.00, 5.00]	5.00 [3.00, 6.00]	2.260	0.024
Neutropenia	Yes	71 (11.53%)	53 (9.46%)	18 (32.14%)	25.676	<0.001
	No	545 (88.47%)	507 (90.54%)	38 (67.86%)		
Hypoalbuminemia (<35 g/L)	Yes	159 (25.81%)	124 (22.14%)	35 (62.50%)	43.300	<0.001
	No	457 (74.19%)	436 (77.86%)	21 (37.50%)		
Lymphocytopenia	Yes	174 (28.25%)	137 (24.46%)	37 (66.07%)	43.483	<0.001
	No	442 (71.75%)	423 (75.54%)	19 (33.93%)		
ICI type	PD-1 inhibitor	399 (64.77%)	364 (65.00%)	35 (62.50%)	0.139	0.709
	PD-L1 inhibitor	217 (35.23%)	196 (35.00%)	21 (37.50%)		
Chemotherapy regimen	Platinum doublet	493 (80.03%)	448 (80.00%)	45 (80.36%)	0.004	0.949
	Single-agent	123 (19.97%)	112 (20.00%)	11 (19.64%)		

Treatment line	First-line	426 (69.16%)	392 (70.00%)	34 (60.71%)	2.058	0.151
	≥Second-line	190 (30.84%)	168 (30.00%)	22 (39.29%)		
Corticosteroid use	Yes	134 (21.75%)	112 (20.00%)	22 (39.29%)	11.124	<0.001
	No	482 (78.25%)	448 (80.00%)	34 (60.71%)		
Prior radiotherapy	Yes	98 (15.91%)	84 (15.00%)	14 (25.00%)	3.805	0.051
	No	518 (84.09%)	476 (85.00%)	42 (75.00%)		
Prior surgery	Yes	122 (19.81%)	112 (20.00%)	10 (17.86%)	0.147	0.701
	No	494 (80.19%)	448 (80.00%)	46 (82.14%)		
Prophylactic G-CSF use	Yes	210 (34.09%)	196 (35.00%)	14 (25.00%)	2.266	0.132
	No	406 (65.91%)	364 (65.00%)	42 (75.00%)		
Prophylactic antifungal use	Yes	64 (10.39%)	56 (10.00%)	8 (14.29%)	1.004	0.316
	No	552 (89.61%)	504 (90.00%)	48 (85.71%)		

Note: BMI, body mass index; ECOG, Eastern Cooperative Oncology Group; PD-L1, programmed death-ligand 1; COPD, chronic obstructive pulmonary disease; CRP, C-reactive protein; G-CSF, granulocyte colony-stimulating factor; WBC, white blood cell; ICI, immune checkpoint inhibitor; SCLC, small cell lung cancer.

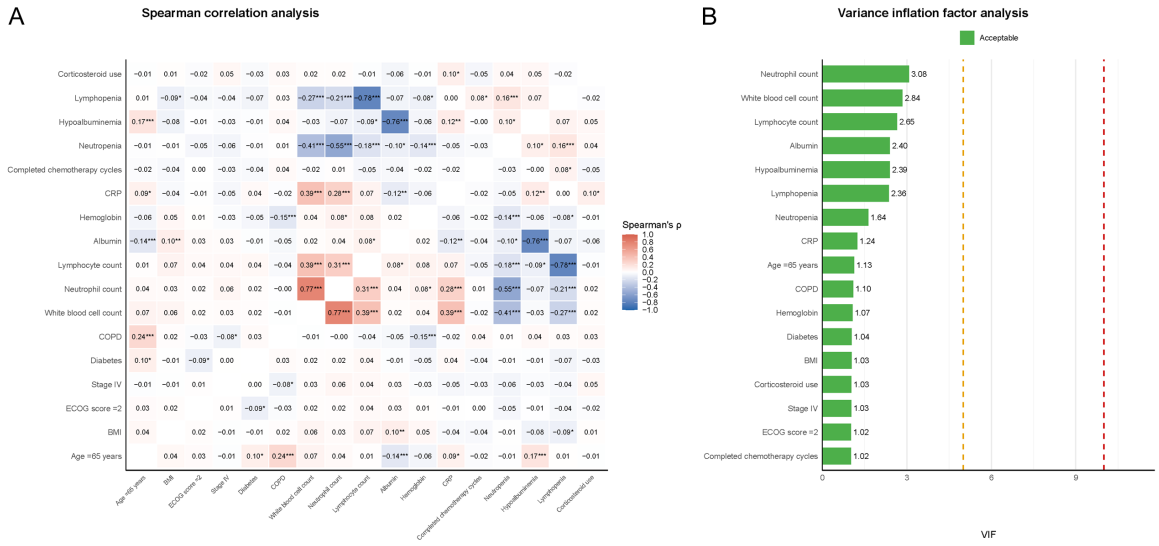

Figure 2. Collinearity assessment of candidate variables. A. Heatmap for spearman correlation coefficients among the 17 candidate variables included in pre-modeling collinearity assessment. Color intensity represents Spearman's rank correlation coefficient (ρ), ranging from -1 to 1, with blue indicating negative correlation, white indicating no correlation, and red indicating positive correlation. B. Variance inflation factor (VIF) plot of the same candidate variables. The dashed reference line indicates VIF=5, and values below this threshold were considered to indicate no substantial multicollinearity. Note: BMI, body mass index; ECOG, Eastern Cooperative Oncology Group; COPD, chronic obstructive pulmonary disease; CRP, C-reactive protein; VIF, variance inflation factor; r, Spearman correlation coefficient.

were 36.350 g/L (sensitivity 76.8%, specificity 70.0%) for albumin and 0.995×10^9/L (sensitivity 66.1%, specificity 75.5%) for lymphocyte count (**Figure 5B**, **5C**). Hemoglobin had an AUC of 0.746 (95% CI: 0.683-0.809) with a cutoff of 112.500 g/L (sensitivity 91.1%, specificity 52.0%) (**Figure 5D**). Neutrophil count had an AUC of 0.691 (95% CI: 0.616-0.766) with a cutoff of 2.615×10^9/L (sensitivity 55.4%, specificity 77.0%) (**Figure 5A**). Completed chemotherapy cycles had the lowest AUC (AUC=0.590, 95% CI: 0.511-0.670) with a cutoff of 4.5 cycles (sensitivity 55.4%, specificity 61.3%) (**Figure 5F**). The Youden index, likelihood ratios, and predictive values for each predictor are presented in Table S2. Pairwise DeLong comparisons showed that CRP had a significantly higher AUC than neutrophil count (P=0.007), hemoglobin (P=0.047), and completed chemotherapy cycles (P<0.001), but was not significantly different from lymphocyte count (P=0.138) or albumin (P=0.112) (Table S3).

Risk score construction and stratification

A clinical risk scoring system was developed using seven dichotomized variables. Points were assigned by re-fitting the multivariable

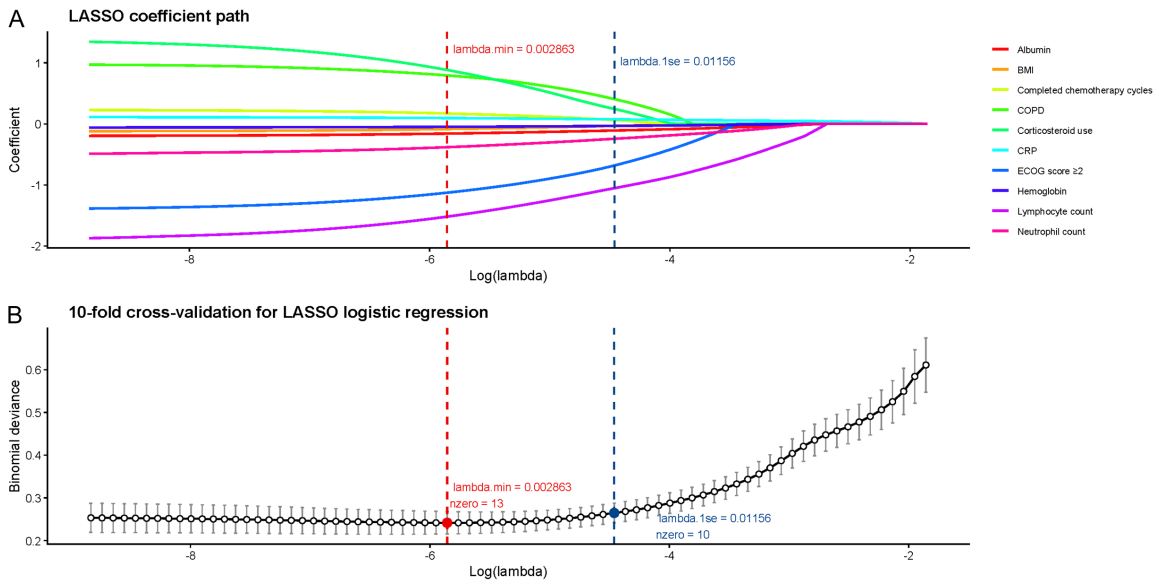

Figure 3. LASSO regression-based candidate variable selection. A. LASSO regression coefficient path plot for the 14 candidate variables entered into LASSO regression after removal of redundant binary indicators. The red dashed line indicates lambda.min and the blue dashed line indicates lambda.1se. B. Ten-fold cross-validation curve for the LASSO logistic regression model using binomial deviance as the optimization criterion. At lambda.min=0.0029, white blood cell count was excluded and 13 variables were retained. At lambda.1se=0.0116, age ≥65 years, stage IV, and diabetes were additionally excluded, leaving 10 variables for subsequent multivariable analysis. Note: LASSO, least absolute shrinkage and selection operator; BMI, body mass index; ECOG, Eastern Cooperative Oncology Group; COPD, chronic obstructive pulmonary disease; CRP, C-reactive protein.

Multivariable logistic regression analysis

Variable	Assignment	OR (95% CI)	P value
Age ≥65 years	1 = ≥65 years; 0 = <65 years	1.165 (0.412-3.268)	0.770
BMI	Continuous, per 1 kg/m² increase	0.899 (0.77-1.037)	0.157
ECOG score ≥2	1 = ECOG score ≥2; 0 = <2	4.18 (1.564-11.736)	0.005
COPD	1 = yes; 0 = no	2.777 (0.876-8.905)	0.081
Neutrophil count	Continuous, per 1 ×10⁹/L increase	0.629 (0.471-0.815)	<0.001
Lymphocyte count	Continuous, per 1 ×10⁹/L increase	0.174 (0.06-0.45)	<0.001
Albumin	Continuous, per 1 g/L increase	0.82 (0.734-0.907)	<0.001
Hemoglobin	Continuous, per 1 g/L increase	0.942 (0.909-0.973)	<0.001
CRP	Continuous, per 1 mg/L increase	1.12 (1.083-1.166)	<0.001
Completed chemotherapy cycles	Continuous, per 1 cycle increase	1.29 (1.017-1.669)	0.042
Corticosteroid use	1 = yes; 0 = no	2.669 (0.909-7.822)	0.071

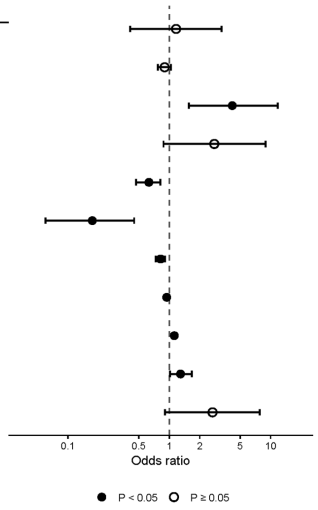

Figure 4. Forest plot of multivariable logistic regression analysis for infection risk factors. The multivariable model included the 10 variables retained by LASSO at lambda.1se, together with age ≥65 years, which was added based on clinical judgment. Odds ratios are shown with 95% confidence intervals. OR >1 indicates increased infection risk, whereas OR <1 indicates reduced infection risk. Note: OR, odds ratio; CI, confidence interval; BMI, body mass index; ECOG, Eastern Cooperative Oncology Group; COPD, chronic obstructive pulmonary disease; CRP, C-reactive protein.

logistic regression model and scaling each regression coefficient (β) relative to the largest absolute β (CRP: β=3.822), multiplied by 3, and rounded to the nearest integer. Variables with a raw score <1 were assigned 1 point. For protective factors (higher values associated with lower infection risk: neutrophil count, lymphocyte count, albumin, and hemoglobin), points

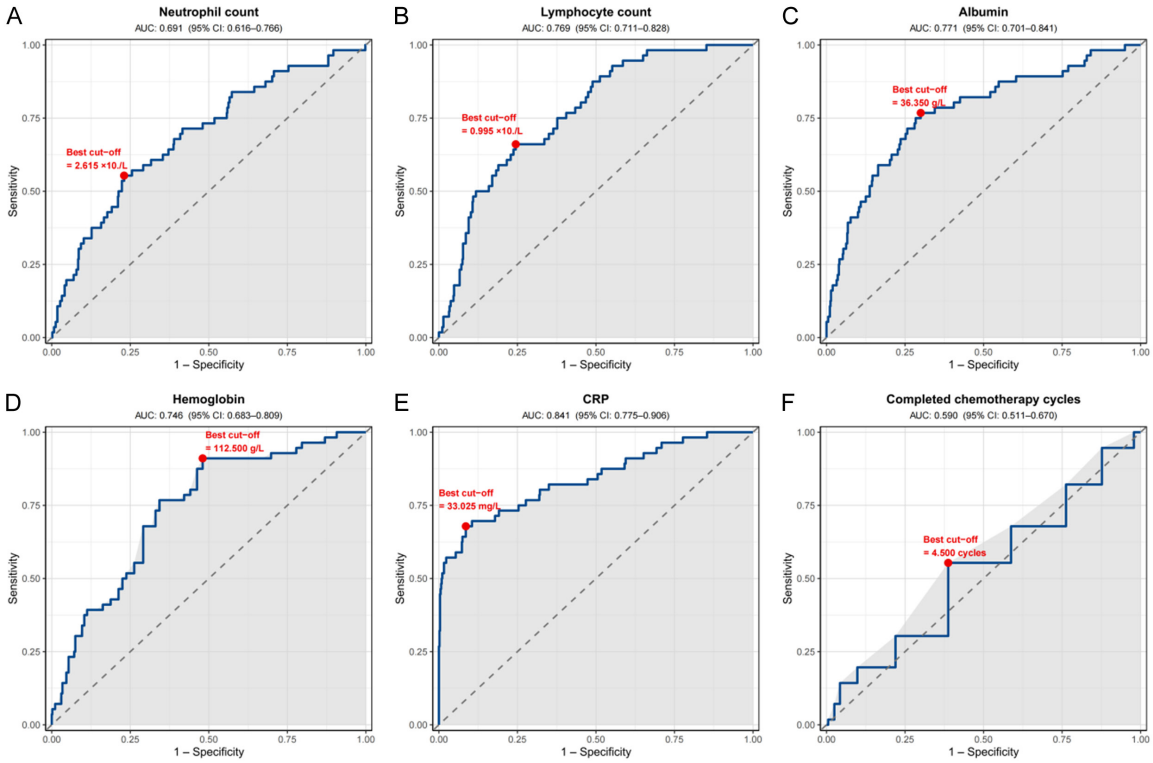

Figure 5. Receiver operating characteristic (ROC) curve analysis of six continuous independent predictors of infection. A. Neutrophil count; B. Lymphocyte count; C. Albumin; D. Hemoglobin; E. C-reactive protein (CRP); F. Completed chemotherapy cycles. ROC curves were constructed using infection status as the binary outcome. The area under the curve (AUC) and 95% confidence interval (CI) were estimated using the DeLong method. Optimal cutoff values were determined by the maximum Youden index, and the red dot indicates the corresponding operating point on each ROC curve. For variables inversely associated with infection risk (neutrophil count, lymphocyte count, albumin, and hemoglobin), direction was standardized during ROC computation so that higher transformed values indicated higher infection risk, while cutoff values are reported in the original clinical units. The x-axis represents 1 - specificity and the y-axis represents sensitivity.

were assigned when the value fell below the ROC-derived cutoff, reflecting increased risk.

The resulting point assignments were: ECOG ≥2 (β=1.553, 1 point), neutrophil count ≤2.615×10^9/L (β=1.311, 1 point), lymphocyte count ≤0.995×10^9/L (β=1.680, 1 point), cumulative completed chemotherapy cycles ≥4.5 (β=0.914, 1 point), albumin ≤36.350 g/L (β=1.993, 2 points), hemoglobin ≤112.500 g/L (β=2.635, 2 points), and CRP ≥33.025 mg/L (β=3.822, 3 points). The total score ranged from 0 to 11 points. Infection incidence increased steadily with higher scores (**Figure 6A**). Stratifying patients into low-risk (0-3), intermediate-risk (4-7), and high-risk (8-11) groups yielded infection rates of 0.27% (1/374), 11.27% (23/204), and 84.21% (32/38), with significant differences across risk strata (**Figure 6B**).

Bootstrap internal validation

Internal validation with 1,000 bootstrap resamples showed that the bootstrap-validated AUC (0.954, 95% CI: 0.926-0.977) closely matched the original AUC (0.954, 95% CI: 0.928-0.979), indicating good internal stability (**Figure 7A**). Using a cutoff of 5.5 points, the bootstrap-validated sensitivity and specificity were 85.7% and 90.1%, respectively, (original: 85.7% and 90.2%). The Brier score was 0.035 (original: 0.036), indicating minimal prediction error. Calibration curve analysis showed excellent alignment between predicted and observed probabilities, with a calibration intercept of 0.000 (95% CI: -0.469 to 0.496), a calibration slope of 1.000 (95% CI: 0.796-1.247), and a non-significant Hosmer-Lemeshow test (χ^2=8.993, P=0.109) (**Figure 7B**).

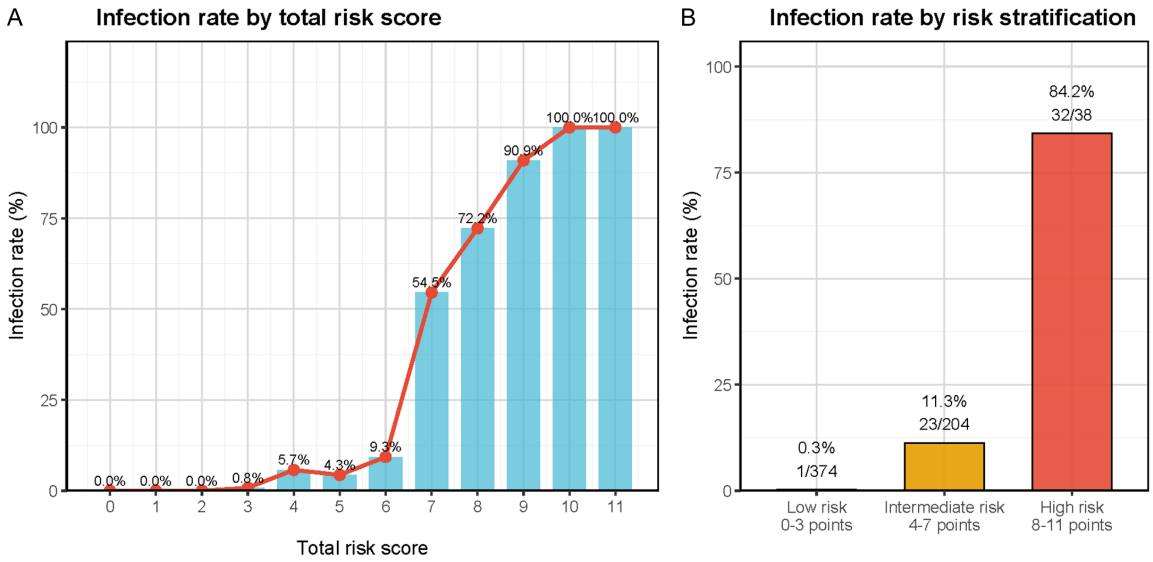

Figure 6. Infection incidence by total risk score and risk stratum. A. Distribution of infection incidence across individual total risk scores; B. Comparison of infection incidence among low-, intermediate-, and high-risk groups. Note: ECOG, Eastern Cooperative Oncology Group; CRP, C-reactive protein.

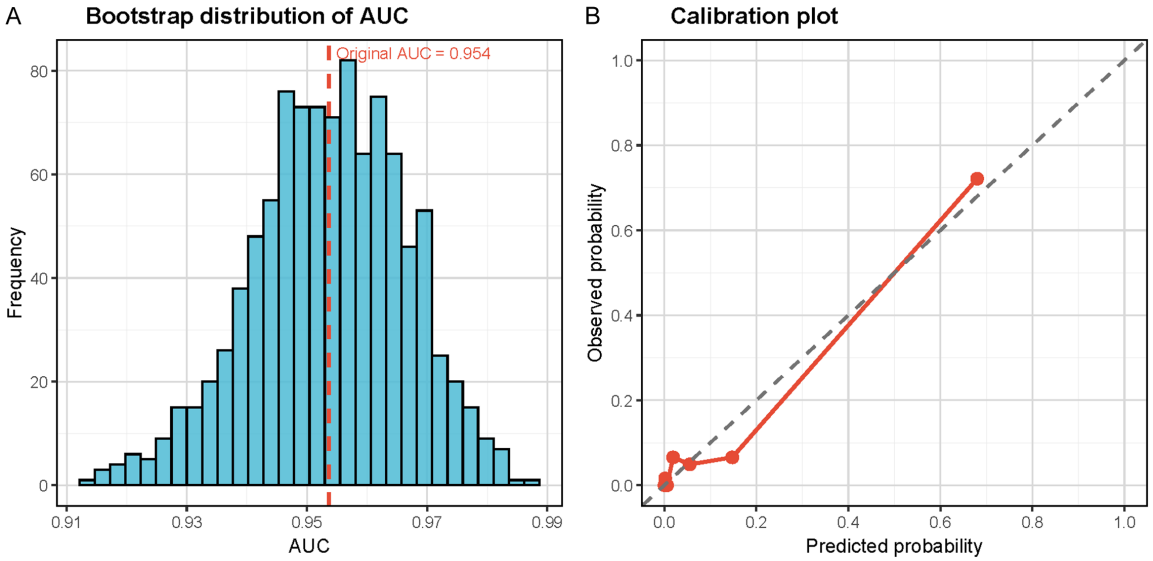

Figure 7. 1,000-iteration bootstrap internal validation results of the risk scoring system. A. Histogram of AUC distribution across 1,000 bootstrap resamples; red dashed line indicates the original AUC value (0.954). B. Calibration curve assessed by decile-based risk grouping; the red line shows the correspondence between mean predicted probability and observed event rate within each decile, and the grey dashed line represents the ideal calibration reference. Calibration intercept =0.000 (95% CI: -0.469 to 0.496), calibration slope =1.000 (95% CI: 0.796-1.247), Hosmer-Lemeshow test χ^2=8.993, P=0.109. Note: AUC, area under the curve; CI, confidence interval; Bootstrap, bootstrap resampling.

Subgroup analysis

To evaluate the generalizability of the risk score across different clinical subgroups, we performed analyses stratified by histological type, ICI class, treatment line, and ECOG status. OR per 1-point increase in score was used to quantify predictive performance.

The score significantly predicted infection risk for squamous cell carcinoma (SCC) with OR=4.563 (95% CI: 2.439-8.534, P<0.001),

Subgroup	Level	Events/N	Rate	OR (95% CI)	P interaction
Pathology type	Squamous cell carcinoma	20/216	9.26%	4.563 (2.439-8.534)	0.387
Pathology type	Adenocarcinoma	25/305	8.20%	2.801 (2.036-3.854)	0.387
Pathology type	Small cell lung cancer	8/75	10.67%	4.854 (1.629-14.470)	0.387
Pathology type	Others	3/20	15.00%	5.608 (0.580-54.274)	0.387
Immunotherapy drug	PD-L1 inhibitor	21/217	9.68%	2.999 (2.009-4.476)	0.394
Immunotherapy drug	PD-1 inhibitor	35/399	8.77%	3.833 (2.592-5.667)	0.394
Treatment line	First-line	34/426	7.98%	3.400 (2.418-4.779)	0.959
Treatment line	Second-line or later	22/190	11.58%	3.350 (2.121-5.290)	0.959
ECOG status	ECOG 0–1	28/448	6.25%	3.330 (2.352-4.716)	0.893
ECOG status	ECOG ≥2	28/168	16.67%	3.463 (2.198-5.456)	0.893

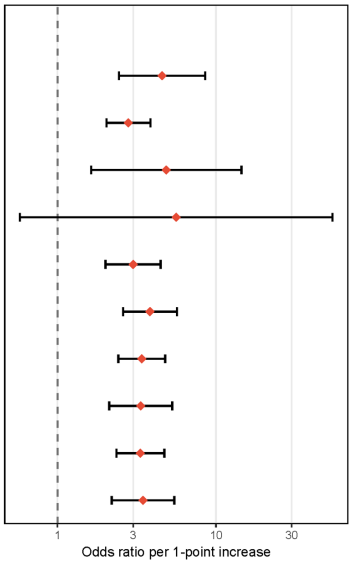

Figure 8. Subgroup analysis forest plot of predictive performance of the risk scoring system across clinical subgroups. Note: OR, odds ratio; CI, confidence interval; ECOG, Eastern Cooperative Oncology Group; PD-1, programmed death-1; PD-L1, programmed death-ligand 1.

adenocarcinoma (ADC) OR=2.801 (95% CI: 2.036-3.854, P<0.001), and SCLC OR=4.854 (95% CI: 1.629-14.470, P=0.005). The "other histology" subgroup did not reach significance, as few events occurred (P=0.137). The interaction test was not significant. By ICI class, PD-L1 inhibitors (OR=2.999, 95% CI: 2.009-4.476, P<0.001) and PD-1 inhibitors (OR=3.833, 95% CI: 2.592-5.667, P<0.001) yielded consistent results (P interaction=0.394). First-line (OR=3.400, 95% CI: 2.418-4.779, P<0.001) and second-line or later (OR=3.350, 95% CI: 2.121-5.290, P<0.001) subgroups were similarly consistent (P interaction=0.959). ECOG 0-1 (OR=3.330, 95% CI: 2.352-4.716, P<0.001) and ECOG ≥2 (OR=3.463, 95% CI: 2.198-5.456, P<0.001) showed stable predictive performance (P interaction=0.893). These findings indicate that the score performs consistently across clinical subgroups (**Figure 8**).

Sensitivity analysis

To assess the potential impact of overfitting related to the number of candidate variables, the full 11-variable model was compared with a reduced model comprising only the seven predictors included in the risk score. Model discrimination was assessed using ROC analysis, and AUCs were compared using the DeLong test.

The reduced model achieved an AUC of 0.976 (95% CI: 0.962-0.989), which did not differ significantly from the full model (AUC=0.978, 95% CI: 0.965-0.991; DeLong P=0.265). AIC was slightly lower for the reduced model (146.17 vs. 146.96), and sensitivity/specificity were comparable (92.9%/92.5% vs. 92.9%/94.3%), suggesting that the seven predictors captured the majority of predictive information.

Three additional sensitivity analyses were performed. First, the simplified risk score (AUC=0.954, sensitivity 85.7%, specificity 90.2%) was comparable to that of the full logistic regression model (AUC=0.958, sensitivity 83.9%, specificity 96.1%; DeLong P=0.121). In contrast, excluding CRP reduced AUC to 0.897 (sensitivity 92.9%, specificity 72.3%; DeLong P<0.001 vs. the full score) (**Figure 9A**, **9B**). Second, reclassification analysis confirmed that including CRP significantly improved model performance (NRI=1.046, 95% CI: 0.801-1.289; IDI=0.061, 95% CI: 0.027-0.099) (**Figure 9C**). Third, an alternative stratification (low-risk 0-4, intermediate-risk 5-7, high-risk 8-11) yielded infection rates of 1.1% (5/444), 14.2% (19/134), and 84.2% (32/38), consistent with the original scheme (**Figure 9D**).

Discussion

In this study, we analyzed 616 lung cancer inpatients treated with ICIs in combination with che-

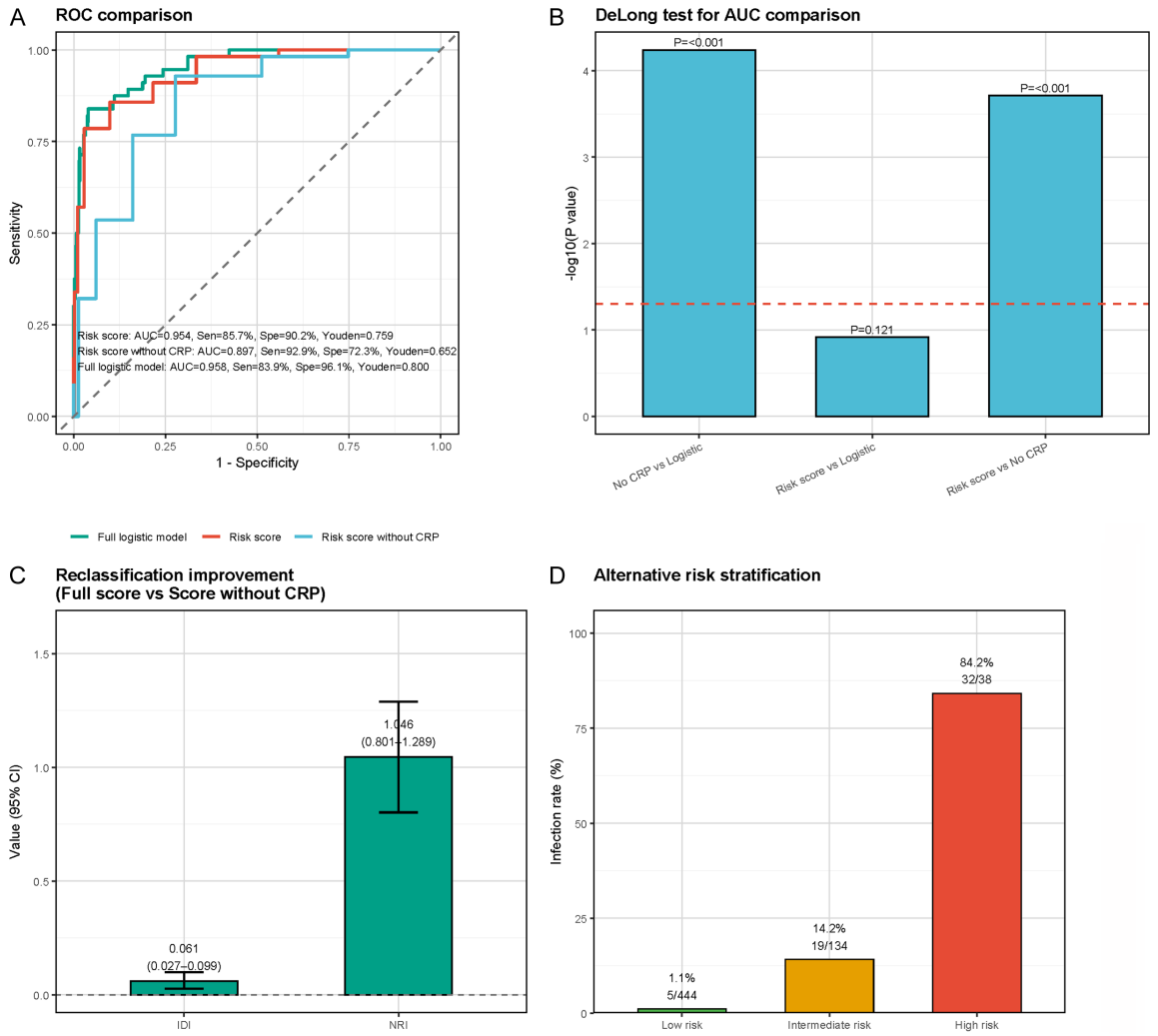

Figure 9. Sensitivity analyses of the risk scoring system. A. ROC curve comparison among three predictive models; B. Statistical significance of AUC differences between models assessed by DeLong test; C. NRI and IDI reclassification improvement results comparing the full score vs. the score with CRP excluded; D. Infection incidence corresponding to the alternative risk stratification scheme. Note: ROC, receiver operating characteristic; AUC, area under the curve; CRP, C-reactive protein; NRI, net reclassification improvement; IDI, integrated discrimination improvement; DeLong, DeLong test.

motherapy. The overall infection rate was 9.09%, consistent with previously published data. Infection rates in lung cancer patients undergoing chemotherapy vary greatly (5%-20%), largely due to heterogeneity in baseline characteristics, diagnostic criteria, observation periods, and treatment regimens. For example, Burns et al. [17] reported an infection incidence up to 45.7% in NSCLC patients treated with pembrolizumab. In that cohort, infected patients experienced significantly higher rates of emergency department visits, hospitalizations, and ICU admissions, and had significantly shorter median OS. In the present study, infection events were defined according to standardized clinical and microbiological criteria based on well-established diagnostic guidelines, and the study population was relatively homogeneous (all receiving ICI-based chemotherapy). These factors likely contributed to the internal consistency of the reported infection incidence. Moreover, this study focused on infections occurring during treatment, which seem to better reflect real-world risk under immunochemotherapy. As far as we know, few studies have developed a practical risk scoring

system for infections in patients receiving immunochemotherapy, highlighting the novelty and clinical relevance of our work.

Chemotherapy impairs myelopoiesis, leading to neutropenia and lymphocytopenia, which compromise host immunity [18]. ICIs may additionally induce immune dysregulation [12]. However, the additive effect of ICIs and chemotherapy on infection risk remains unclear [19]. In a comparative study, infection rates in patients receiving ICI plus chemotherapy and chemotherapy alone were similar (15% vs. 22%, P=0.1) [20], suggesting the infection hazard is likely driven by myelosuppression and the concomitant use of immunosuppressive agents rather than by ICIs themselves. Supporting this, Bavaro et al. [21] identified corticosteroid use for irAE management as an independent predictor of infection in NSCLC patients receiving ICIs (adjusted HR=2.65, P=0.009), and neutropenia was also significantly associated with infection. These results shed light on the complex nature of infection risk associated with combination therapy and highlight the need for careful management of corticosteroids.

The most common infection site was the lung (64.29%). This observation aligns with the pathophysiology of lung cancer, in which the primary tumor disrupts airway architecture, bronchial obstruction impairs clearance of secretions, chemotherapy damages the mucosal barrier, and the risk of aspiration is increased. Bloodstream infection (14.29%) was the second most common site, indicating a significant burden of severe infection that requires close clinical attention. Urinary tract infection occurred in 10.71% of cases, likely due to the use of indwelling catheters and immune compromise. Gram-negative pathogens comprised 50.00% of isolates with *K. Pneumonia*, *Pseudomonas aeruginosa*, and E. *Coli* being predominant. Zheng et al. [22] reported a predominance of Gram-negative microorganisms in pulmonary infections among patients receiving chemotherapy for advanced lung cancer (64.29%), which is consistent with our study. This pattern likely reflects the predictable ecology of microorganisms in these patients. Gram-positive organisms comprised 25.00%, while fungi were isolated in 14.29% of cases. Previous studies indicate that fungal pulmonary infections occurs in about 14.00% of lung cancer patients, with corticosteroid use (OR=4.44) and receipt of two or more chemotherapy cycles (OR=2.87) identified as independent risk factors [23]. These findings underscore the importance of incorporating antifungal prophylaxis or monitoring in treatment planning.

Drug-resistant organisms, including CRKP, MRSA, and CRPA, were detected among culture-positive cases. However, given the limited number of isolates, these proportions should be interpreted with caution in relation to local surveillance data; they do not allow general conclusions on the overall resistance. In the era of immunotherapy, the spectrum of opportunistic infections in lung cancer patients is evolving [24]. Infection risk is heightened in patients with absolute lymphocyte count <500/μL or those receiving prolonged use of corticosteroids (≥20 mg prednisone equivalent/day for ≥4 weeks). Drug-resistant infections are more difficult to treat, often prolong hospitalization and may compromise treatment efficacy [25]. The emergence of these organisms is greatly driven by repeated hospital admissions and recurrent exposure to broad-spectrum antibiotics. Therefore, local antimicrobial resistance surveillance should guide empirical antibiotic selection to prevent further escalation of resistance [26].

Infection events clustered during chemotherapy cycles 3-4 of (35.71%), consistent with the timeline of cumulative treatment-related myelosuppression. Repeated cycles of chemotherapy progressively impair bone marrow reserve, resulting in sustained reductions in neutrophil and lymphocyte counts. Therefore, enhanced surveillance and timely preventive interventions during this treatment period may be particularly important for reducing infection-related complications.

Multivariable analysis identified seven independent predictors of infection, among which ECOG score ≥2 (OR=3.993) emerged as a strong predictor, reflecting not only diminished physiological reserve and impaired immune competence but also poor nutritional status and overall functional decline. Burns et al. [17] and Bachlitzanaki et al. [27] similarly reported that poor performance status was independently associated with infection or infection-related adverse outcome (OR/HR=5.79, P=0.004), corroborating our findings and highlighting the

important role functional status in infection susceptibility among patients with lung cancer. Elevated CRP was also identified as an independent risk factor (OR=1.114), suggesting a close relationship between systemic inflammation and infection susceptibility. As an acute-phase reactant, CRP reflects inflammatory activation in tumor microenvironment. Importantly, elevated CRP levels before the onset of clinically apparent infection may represent an early warning signal of increased vulnerability to infectious complications. Nardone et al. [28] reported that elevated CRP levels were associated with poor outcomes in patients with metastatic NSCLC receiving PD-1 and PD-L1 blockade. This is consistent with the concept that systemic inflammation adversely affects both anti-tumor and antimicrobial immune responses. Nevertheless, distinguishing tumor-related inflammation from pre-infectious inflammatory activation remains challenging in clinical practice, and dynamic monitoring CRP may be more informative than a single assessment. The number of completed chemotherapy cycles was also independently associated with infection risk (OR=1.317), suggesting a dose-response relationship between infection susceptibility and duration of chemotherapy. With increasing treatment cycles, prolonged myelosuppression, progressive mucosal barrier injury, and cumulative immune dysfunction may collectively contribute to a higher risk of infection. Previous predictive models for pulmonary infection in lung cancer patients undergoing chemoradiotherapy have also identified chemotherapy cycle count as an important predictor [29], indicating that cumulative treatment exposure may represent a common risk factor across different therapeutic settings.

Higher neutrophil count was identified as an independent protective factor for infection (OR=0.636), underscoring its significant role for innate immunity in host defense against pathogens. Previous studies have identified neutropenia as an independent predictor of infection in patients treated with ICI-based chemotherapy [20], consistent with our findings. Lymphocyte count demonstrated the strongest protective association among all predictors (OR=0.174), emphasizing the importance of adaptive immunity. Although ICIs enhance anti-tumor immunity by restoring T-cell activity through blockade of the PD-1/PD-L1 pathway, the immunostimulatory effects of ICIs may be partially offset by chemotherapy-induced lymphocyte depletion. Consequently, lymphocytopenia may remain a major determinant of infection susceptibility despite ongoing immunotherapy. Albumin was also identified as an independent protective factor (OR=0.816). Consistent with our finding, Ding et al. [30] reported that decreased post-chemotherapy albumin levels were independently associated with pulmonary infection after NSCLC resection. It is suggested that anemia appears to increases susceptibility to infections by hindering the metabolism of immune cells and lowering the oxygenation of a tissue. Corticosteroid use approached, but did not reach, statistical significance in the multivariable analysis (P= 0.065). Although this association did not meet the conventional threshold for significance, the observed trend suggests a potential contribution of corticosteroid exposure to infection risk. Guo et al. [31] reported that corticosteroid exposure was an independent risk factor for pulmonary infection in lung cancer patients receiving palliative chemotherapy. Their neural network-based prediction model further demonstrated the importance of corticosteroid use and myelosuppression in infection risk assessment (AUC=0.897).

LASSO regression was employed for variable selection as it effectively reduces overfitting. The optimal penalty parameter (λ) was determined using 10-fold cross-validation. We selected the lambda.1se instead of lambda.min to achieve a more parsimonious model while maintaining adequate predictive performance. Comparable LASSO-logistic models in similar settings have produced internal validation C-indices of 0.900 and AUCs of 0.875-0.897 [29, 30]. In comparison, the AUC of 0.954 achieved by our score suggests some advantage, which may be partly attributable to the inclusion of CRP and immune-related variables reflecting both inflammatory and immune status.

The risk score demonstrated excellent discrimination (AUC=0.954), with predictive performance comparable to that of the full logistic model (AUC=0.958; DeLong P=0.121), while offering substantially greater clinical practicality. Furthermore, the three risk strata corresponded to strikingly different infection rates

(0.27%, 11.27%, 84.21%), providing clinically meaningful risk categorization. The model also exhibited robust internal stability, as evidenced by bootstrap validation and excellent calibration, with good agreement between predicted and observed probabilities. In addition, subgroup analyses demonstrated consistent predictive performance across histological types, ICI classes, treatment lines, and ECOG performance status. Sensitivity analyses further confirmed the importance of CRP, as its exclusion resulted in a significant decline in predictive performance (AUC dropped from 0.954 to 0.897, DeLong P<0.001).

From a clinical perspective, patients in the low-risk group (0-3 points) can be managed with routine surveillance alone; Patients in the intermediate-risk group (4-7 points) may benefit from closer monitoring of infection-related parameters and individualized assessment of preventive measures; Patients in the high risk group (8-11 points) may require intensified preventive strategies, such as optimization of nutritional support, correction of anemia and hypoalbuminemia, careful evaluation of antimicrobial prophylaxis, patient education regarding infection symptoms, and enhanced clinical surveillance. As some components of the score are dynamic and may change during the course of treatment, reassessment before each treatment cycle may allow timely risk reclassification and facilitate personalized infection management.

Limitations

Several limitations should be acknowledged. First, this was a single-center retrospective study conducted at a tertiary hospital in central China. The relatively high proportion of patients with advanced-stage disease (66.40%) may limit the generalizability. The six-year study period also introduces potential temporal confounding of the findings to other populations and healthcare settings. Second, while 616 patients were enrolled, the modest number of infection events (n=56) limited the statistical power of subgroup analyses and precluded external validation. Therefore, the proposed risk score requires validation in prospective, multicenter, independent cohorts before clinical application. Third, infection was defined based on the Chinese Diagnostic Criteria for Nosocomial Infections (2001). Differences between these criteria and internationally adopted standards, such as the CDC/NHSN criteria, may affect cross-study comparability. Fourth, despite multivariable adjustment, residual confounding cannot be completely excluded. Several clinically relevant markers, including IL-6, procalcitonin, the neutrophil-to-lymphocyte ratio, and the platelet-to-lymphocyte ratio, were not routinely available and therefore could not be incorporated into the analysis. Similarly, detailed nutritional assessment data, detailed antibiotic exposure history, gut microbiome composition, and prior infection-related hospitalizations were unavailable, potentially limiting the completeness of risk factor evaluation. Fifth, the events-per-variable ratio of the final multivariable model was approximately 5.1 (56 infection events/11 candidate variables), which is below the conventional 10 events per variable. However, a sensitivity analysis comparing the full 11-variable model with a reduced 7-variable model (EPV=8.0) showed nearly identical AUC values (0.978 vs. 0.976; DeLong P=0.265), suggesting that overfitting from variable proliferation was not a major concern. Nevertheless, LASSO regression and bootstrap internal validation cannot replace external validation, and the proposed risk score should be regarded as exploratory until validated in prospective multicenter cohorts.

Conclusion

This study characterizes the epidemiology of infections among lung cancer patients receiving ICI-based chemotherapy. Infections were clinically significant, occurred predominantly in the lungs, and were associated with a notable proportion of drug-resistant pathogens, underscoring the need for heightened clinical attention. A risk scoring system was developed based on seven easily available clinical variables and showed good discriminatory performance and internal stability. These findings suggest that the proposed tool may facilitate early identification of high-risk patients and support individualized preventive and monitoring strategies in routine clinical practice. Future studies should focus on external validation of this scoring system in independent multicenter cohorts and explore whether dynamic, score-guided infection prevention strategies can improve treatment tolerance, reduce infectious

complications, and ultimately enhance clinical outcomes in patients receiving immunochemotherapy.

Disclosure of conflict of interest

None.

Address correspondence to: Dingfang Pi, Department of Infectious Diseases, The First Affiliated Hospital of Yangtze University/The First People's Hospital of Jingzhou, No. 55, Jianghan Road, Shashi District, Jingzhou 434000, Hubei, China. E-mail: 18163139085@163.com; Yanxia Zhang, Pulmonary and Critical Care Medicine, The First Affiliated Hospital of Yangtze University/The First People's Hospital of Jingzhou, No. 55, Jianghan Road, Shashi District, Jingzhou 434000, Hubei, China. E-mail: fish123sky@163.com

References

[1] Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I and Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2024; 74: 229-263.

[2] Han B, Zheng R, Zeng H, Wang S, Sun K, Chen R, Li L, Wei W and He J. Cancer incidence and mortality in China, 2022. J Natl Cancer Cent 2024; 4: 47-53.

[3] Paz-Ares L, Dvorkin M, Chen Y, Reinmuth N, Hotta K, Trukhin D, Statsenko G, Hochmair MJ, Özgüroğlu M, Ji JH, Voitko O, Poltoratskiy A, Ponce S, Verderame F, Havel L, Bondarenko I, Kazarnowicz A, Losonczy G, Conev NV, Armstrong J, Byrne N, Shire N, Jiang H and Goldman JW. Durvalumab plus platinum-etoposide versus platinum-etoposide in first-line treatment of extensive-stage small-cell lung cancer (CASPIAN): a randomised, controlled, open-label, phase 3 trial. Lancet 2019; 394: 1929-1939.

[4] Parmar A, Lu B, Luo J and W Chan KK. Real-world comparative effectiveness and safety of pembrolizumab for PD-L1≥50% metastatic non-small-cell lung cancer. Future Oncol 2024; 20: 2879-2888.

[5] Xiong A, Wang L, Chen J, Wu L, Liu B, Yao J, Zhong H, Li J, Cheng Y, Sun Y, Ge H, Yao J, Shi Q, Zhou M, Chen B, Han Z, Wang J, Bu Q, Zhao Y, Chen J, Nie L, Li G, Li X, Yu X, Ji Y, Sun D, Ai X, Chu Q, Lin Y, Hao J, Huang D, Zhou C, Shan J, Yang H, Liu X, Wang J, Shang Y, Mei X, Yang J, Lu D, Hu M, Wang ZM, Li B, Xia M and Zhou C. Ivonescimab versus pembrolizumab for PD-L1-positive non-small cell lung cancer (HARMONi-2): a randomised, double-blind, phase 3 study in China. Lancet 2025; 405: 839-849.

[6] Roussot N, Kaderbhai C and Ghiringhelli F. Targeting immune checkpoint inhibitors for non-small-cell lung cancer: beyond PD-1/PD-L1 monoclonal antibodies. Cancers (Basel) 2025; 17: 906.

[7] Hao Y, Zhang X and Yu L. Immune checkpoint inhibitor-related pneumonitis in non-small cell lung cancer: a review. Front Oncol 2022; 12: 911906.

[8] Jovanoski N, Kaur G, Shukla H, Chana N and Kharawala S. Economic burden of advanced non-small cell lung cancer (NSCLC): a systematic literature review. J Med Econ 2026; 29: 433-454.

[9] GBD 2023 Lower Respiratory Infections and Antimicrobial Resistance Collaborators. Global burden of lower respiratory infections and aetiologies, 1990-2023: a systematic analysis for the Global Burden of Disease Study 2023. Lancet Infect Dis 2026; 26: 343-361.

[10] Schneider SM, Fan C, Wang Y, Jenq RR and Watowich SS. Microbiota and immune-related adverse events in cancer immunotherapy. Nat Rev Cancer 2026.

[11] Yin J, Wu Y, Yang X, Gan L and Xue J. Checkpoint inhibitor pneumonitis induced by anti-PD-1/PD-L1 therapy in non-small-cell lung cancer: occurrence and mechanism. Front Immunol 2022; 13: 830631.

[12] Chen J, Liu Y, Cai H and Zheng W. Risk factors for pulmonary infection in patients with non-small cell lung cancer: a meta-analysis. BMC Pulm Med 2024; 24: 353.

[13] Bavaro DF, Pizzutilo P, Catino A, Signorile F, Pesola F, Di Gennaro F, Cassiano S, Marech I, Lamorgese V, Angarano G, Monno L, Saracino A and Galetta D. Incidence of infections and predictors of mortality during checkpoint inhibitor immunotherapy in patients with advanced lung cancer: a retrospective cohort study. Open Forum Infect Dis 2021; 8: ofab187.

[14] Horan TC, Andrus M and Dudeck MA. CDC/NHSN surveillance definition of health care-associated infection and criteria for specific types of infections in the acute care setting. Am J Infect Control 2008; 36: 309-332.

[15] Connelly JA and Walkovich K. Diagnosis and therapeutic decision-making for the neutropenic patient. Hematology Am Soc Hematol Educ Program 2021; 2021: 492-503.

[16] Chen C, Zhang Y, Zhao X, Tao M, Yan W and Fu Y. Hypoalbuminemia - an indicator of the severity and prognosis of COVID-19 patients: a multicentre retrospective analysis. Infect Drug Resist 2021; 14: 3699-3710.

[17] Burns EA, Gee K, Kieser RB, Xu J, Zhang Y, Crenshaw A, Muhsen IN, Mylavarapu C, Esmail

A, Shah S, Umoru G, Sun K, Guerrero C, Gong Z, Heyne K, Singh M, Zhang J, Bernicker EH and Abdelrahim M. Impact of infections in patients receiving pembrolizumab-based therapies for non-small cell lung cancer. Cancers (Basel) 2022; 15: 81.

[18] Sharma A, Jasrotia S and Kumar A. Effects of chemotherapy on the immune system: implications for cancer treatment and patient outcomes. Naunyn Schmiedebergs Arch Pharmacol 2024; 397: 2551-2566.

[19] Petrelli F, Morelli AM, Luciani A, Ghidini A and Solinas C. Risk of infection with immune checkpoint inhibitors: a systematic review and meta-analysis. Target Oncol 2021; 16: 553-568.

[20] Malek AE, Khalil M, Hachem R, Chaftari AM, Fares J, Jiang Y, Kontoyiannis DP, Fossella F, Chaftari P, Mulanovich VE, Viola GM and Raad II. Impact of checkpoint inhibitor immunotherapy, primarily pembrolizumab, on infection risk in patients with advanced lung cancer: a comparative retrospective cohort study. Clin Infect Dis 2021; 73: e2697-e2704.

[21] Bavaro DF, Diella L, Pizzutilo P, Catino A, Signorile F, Pesola F, Belati A, Marech I, Garrisi V, Lamorgese N, Di Gennaro F, Saracino A and Galetta D. Incidence and predictors of infections in patients with advanced non-small cell lung cancer treated with checkpoint inhibitor immunotherapies: a monocentric retrospective cohort study. Scand J Immunol 2023; 98: e13303.

[22] Zheng X, Lan H, Hu Y and Tian P. Effect of interleukin-6 and interleukin-8 levels on pathogenic bacteria types in patients with advanced lung cancer and pulmonary infection during chemotherapy. Am J Transl Res 2025; 17: 4723-4732.

[23] Deng L, Yu R and Chen B. A meta-analysis on the incidence and associated factors of pulmonary fungal infections in lung cancer patients: evidence predominantly from Chinese studies in the Chinese and English literature. Thorac Cancer 2026; 17: e70258.

[24] Qi W, Tian L, Li Z, Xu J and Wang T. Pathogen spectrum and management strategies for opportunistic infections in lung cancer in the immunotherapy era: recent advances from fungi to mycobacteria. Am J Cancer Res 2025; 15: 5140-5167.

[25] Kalın G, Alp E, Chouaikhi A and Roger C. Antimicrobial multidrug resistance: clinical implications for infection management in critically Ill patients. Microorganisms 2023; 11: 2575.

[26] Elsorady KE, Hassan RA, Abdelhamid DH and Abd El-Mohsen M. Antibiotics use and its association with multi-drug resistance in a tertiary care geriatrics hospital in Egypt. J Infect Dev Ctries 2022; 16: 1860-1869.

[27] Bachlitzanaki M, Aletras G, Bachlitzanaki E, Messaritakis I, Koukias S, Koulouridi A, Bachlitzanakis E, Kaloeidi E, Vakonaki E, Kontopodis E, Androulakis N, Chamilos G, Mavroudis D, Ioannou P and Kofteridis D. Evaluation of febrile neutropenia in hospitalized patients with neoplasia undergoing chemotherapy. Microorganisms 2023; 11: 2547.

[28] Nardone V, Giannicola R, Bianco G, Giannarelli D, Tini P, Pastina P, Falzea AC, Macheda S, Caraglia M, Luce A, Zappavigna S, Mutti L, Pirtoli L, Giordano A and Correale P. Inflammatory markers and procalcitonin predict the outcome of metastatic non-small-cell-lung-cancer patients receiving PD-1/PD-L1 immune-checkpoint blockade. Front Oncol 2021; 11: 684110.

[29] Huang Y and Han G. Predictive nomogram for risk of pulmonary infection in lung cancer patients undergoing radiochemotherapy: development and performance evaluation. Am J Cancer Res 2025; 15: 781-796.

[30] Ding Z, Wang X, Jiang S and Liu J. Risk factors for postoperative pulmonary infection in patients with non-small cell lung cancer: analysis based on regression models and construction of a nomogram prediction model. Am J Transl Res 2023; 15: 3375-3384.

[31] Guo W, Gao G, Dai J and Sun Q. Prediction of lung infection during palliative chemotherapy of lung cancer based on artificial neural network. Comput Math Methods Med 2022; 2022: 4312117.

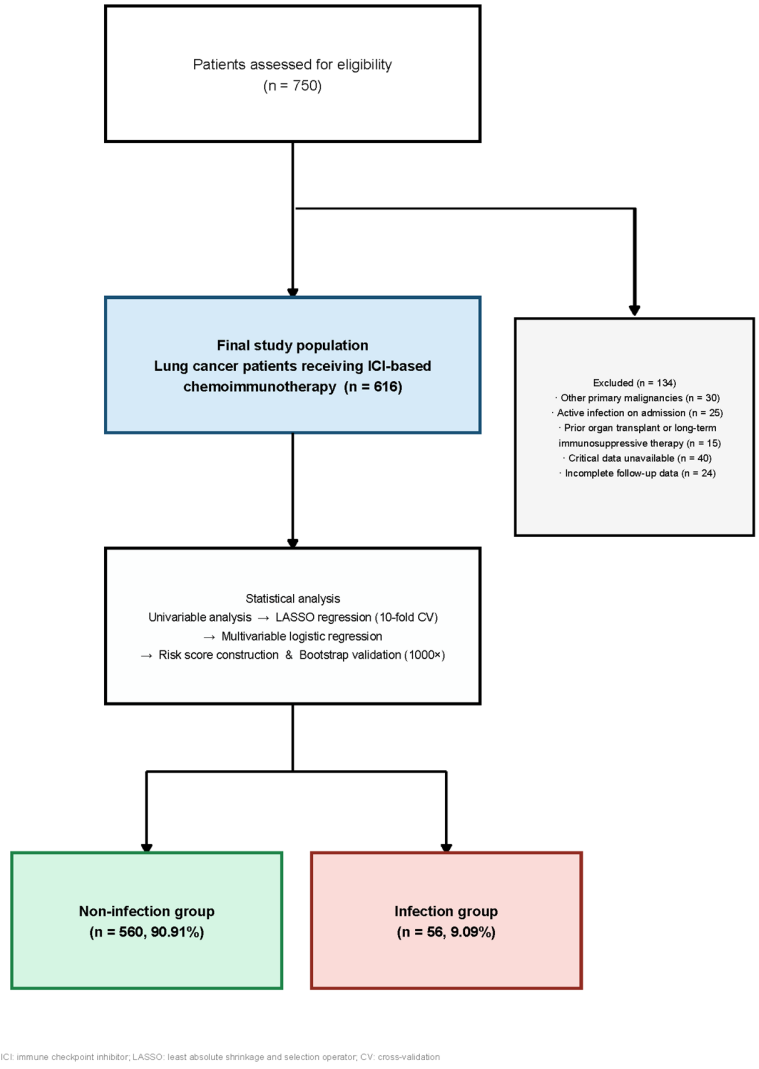

Figure S1. Patient selection flowchart.

Table S1. Normality test results of continuous variables (Shapiro-Wilk test)

Variable	W	P	Distribution
Age (years)	0.9882	<0.001	Non-normal
BMI (kg/m^2)	0.9988	0.9576	Normal
WBC count (×10^9/L)	0.9966	0.2129	Normal
Neutrophil count (×10^9/L)	0.9911	<0.001	Non-normal
Lymphocyte count (×10^9/L)	0.9967	0.2281	Normal
Albumin (g/L)	0.9954	0.066	Normal
Hemoglobin (g/L)	0.9981	0.7163	Normal
CRP (mg/L)	0.8112	<0.001	Non-normal
Cumulative completed chemotherapy cycles	0.9673	<0.001	Non-normal

Note: BMI, body mass index; WBC, white blood cell; CRP, C-reactive protein. Distribution was determined based on P<0.05 indicating non-normal distribution. Normally distributed variables are reported as mean ± SD; non-normally distributed variables are reported as median (IQR).

Figure S2. Study overview infographic. The infographic presents the overall framework and key findings of this study from top to bottom. ① Enrollment and grouping: A total of 616 lung cancer inpatients receiving ICI-based chemotherapy were enrolled; 560 (90.91%) had no infection and 56 (9.09%) developed infection. ② Study timeline and

infection timing: The study period was January 2020 to December 2025; infection events clustered during chemotherapy cycles 3-4 (35.71%). ③ Data collection domains: Six domains were covered - demographics, comorbidities, treatment-related factors, laboratory parameters, infection site, and pathogen profile. ④ Statistical analysis pathway: Univariate analysis → LASSO regression (10-fold cross-validation) → multicollinearity assessment → multivariable logistic regression → ROC curve analysis → 1,000-iteration bootstrap internal validation, which were used to construct and evaluate the risk scoring system. ⑤ Main results: A clinical risk score (0-11 points) was constructed from seven independent variables; infection rates in the low- (0-3), intermediate- (4-7), and high-risk (8-11) groups were 0.27%, 11.27%, and 84.21%, respectively; model AUC was 0.954 (95% CI: 0.928-0.979). Note: ICI, immune checkpoint inhibitor; LASSO, least absolute shrinkage and selection operator; ROC, receiver operating characteristic; AUC, area under the curve; CI, confidence interval; ECOG, Eastern Cooperative Oncology Group; CRP, C-reactive protein.

Table S2. Receiver operating characteristic (ROC) parameters of six continuous predictors for infection

Variable_code	Variable	Risk_ direction	AUC	CI_ lower	CI_ upper	Cutoff	Risk_rule	Sensitivity	Specificity
neutrophil	Neutrophil count	lower	0.691	0.616	0.766	2.615	Neutrophil count ≤2.615×10^9/L	55.4	77
lymphocyte	Lymphocyte count	lower	0.769	0.711	0.828	0.995	Lymphocyte count ≤0.995×10^9/L	66.1	75.5
albumin	Albumin	lower	0.771	0.701	0.841	36.35	Albumin ≤36.350 g/L	76.8	70
hemoglobin	Hemoglobin	lower	0.746	0.683	0.809	112.5	Hemoglobin ≤112.500 g/L	91.1	52
crp	CRP	higher	0.841	0.775	0.906	33.025	CRP ≥33.025 mg/L	67.9	91.4
chemo_cycles	Completed chemotherapy cycles	higher	0.59	0.511	0.67	4.5	Completed chemotherapy cycles ≥4.500 cycles	55.4	61.3

Note: AUC, area under the curve; CI, confidence interval; LR+, positive likelihood ratio; LR-, negative likelihood ratio; PPV, positive predictive value; NPV, negative predictive value; CRP, C-reactive protein. Cutoff values were determined by maximizing the Youden index. PPV and NPV were calculated based on the cohort prevalence of 9.09% (56/616).

Table S3. Pairwise comparison of AUC values using the DeLong test

Variable_1	Variable_2	AUC_1	AUC_2	AUC_diff	Z	P_label
Neutrophil count	Lymphocyte count	0.691	0.769	0.078	-1.804	0.071
Neutrophil count	Albumin	0.691	0.771	0.08	-1.572	0.116
Neutrophil count	Hemoglobin	0.691	0.746	0.055	-1.122	0.262
Neutrophil count	CRP	0.691	0.841	0.15	-2.687	0.007
Neutrophil count	Completed chemotherapy cycles	0.691	0.59	0.101	1.715	0.086
Lymphocyte count	Albumin	0.769	0.771	0.002	-0.034	0.973
Lymphocyte count	Hemoglobin	0.769	0.746	0.024	0.55	0.583
Lymphocyte count	CRP	0.769	0.841	0.071	-1.484	0.138
Lymphocyte count	Completed chemotherapy cycles	0.769	0.59	0.179	3.497	<0.001
Albumin	Hemoglobin	0.771	0.746	0.025	0.504	0.614
Albumin	CRP	0.771	0.841	0.07	-1.589	0.112
Albumin	Completed chemotherapy cycles	0.771	0.59	0.181	3.181	0.001
Hemoglobin	CRP	0.746	0.841	0.095	-1.99	0.047
Hemoglobin	Completed chemotherapy cycles	0.746	0.59	0.155	2.946	0.003
CRP	Completed chemotherapy cycles	0.841	0.59	0.25	4.931	<0.001

Note: AUC, area under the curve; AUC_diff, AUC difference; Z, Z statistic; CRP, C-reactive protein. *P* values were calculated using the DeLong test for paired receiver operating characteristic curves.