

Development and Validation of a Survival Prediction Model for Lung Squamous Cell Carcinoma

Yoshio Ichihashi

Osaka Medical and Pharmaceutical University

Teruyoshi Amagai

amagait@yahoo.co.jp

Jikei University of Health Care Sciences

Shinichi Nakatsuka

Yao Tokushukai General Hospital

Kiyoaki Uryuu

Yao Tokushukai General Hospital

Hiromasa Harada

Yao Tokushukai General Hospital

Kaoru Ochi

Hokusetsu General Hospital

Ayako Tsunou

Kitauwa Hospital

Hiroyo Miyata

Kindai University Hospital

Kiyoshi Sato

Osaka Medical and Pharmaceutical University

Nobuharu Hanaoka

Osaka Medical and Pharmaceutical University

Takahiro Katsumata

Osaka Medical and Pharmaceutical University

Article

Keywords: Lung Squamous Cell Carcinoma, Prediction model, advanced lung cancer inflammation index (ALI), Survival time

Posted Date: July 10th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10041031/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Lung cancer remains the leading cause of cancer-related mortality worldwide. Despite curative surgery, postoperative survival among patients with lung squamous cell carcinoma (LUSC) varies widely, and reliable methods to predict individual survival duration remain limited.

Objective

To develop and validate a predictive model for postoperative survival time in patients with LUSC using readily available clinical biomarkers.

Methods

Male patients with LUSC who underwent radical surgery were retrospectively enrolled from two institutions. The development cohort included patients treated at Osaka Medical and Pharmaceutical University Hospital (2007–2015) and Yao Tokushukai General Hospital (2015–2018). An independent cohort treated at Osaka Medical and Pharmaceutical University Hospital up to 2019 was used for external validation. Clinical variables, laboratory parameters, inflammatory indices, and pathological findings were collected. Predictive variables were identified using logistic regression analysis. A survival time prediction model was developed using Cox proportional hazards regression. Model discrimination and calibration were assessed using Harrell's concordance index (C-index), calibration plots, bootstrap internal validation, and external validation. Clinical utility was evaluated using decision curve analysis.

Results

A total of 69 male patients were included in the analysis. The advanced lung cancer inflammation index (ALI) and hemoglobin (Hb) were identified as independent predictors of survival. The predictive formula for postoperative survival time was:

$$\text{Survival time (days)} = (6.9 \times \text{ALI}) + (93.4 \times \text{Hb}) - 198.4.$$

ROC analysis identified an optimal ALI cutoff value of 48.04 (AUC = 0.690, $p = 0.003$). The model demonstrated acceptable discrimination and calibration in both internal and external validation cohorts. External validation confirmed reproducibility with a C-index of 0.70 (95% CI, 0.68–0.84). Decision curve analysis indicated a favorable net clinical benefit.

Conclusions

A simple survival prediction model incorporating ALI and Hb may help stratify postoperative risk in patients with LUSC and support individualized treatment strategies and follow-up planning.

Background

Lung cancer has emerging as the leading cause of cancer-related mortality (Zhou et al.2024). According to the Global Cancer Observatory (GLOBOCAN), the number of new lung cancer cases increased from 2.2 million in 2020 to 2.48 million in 2024, while the number of lung cancer deaths remained at 1.8 million (Sung et al. 2021; Zhou et al. 2024). The most common type of lung cancer is non-small-cell lung cancer (NSCLC), which consists of adenocarcinoma, lung squamous cell carcinoma (LUSC), and large cell carcinoma, accounting for 85% of lung cancer (Yu et al. 2021). The five-year overall survival rate for LUSC remains less than 15% (Lu et al. 2019). Identification of patients who are likely to experience short or long survival has clinical utility by helping to minimize over- or under-treatment, potentially leading to improved clinical decision making (Miller et al. 2022). Indicators for predicting survival as 30-day mortality (Razdan et al. 2021; Wang et al. 2023) or survival rate within 5 years using various biomarkers (Chen et al. 2021; Pezzella et al. 2019; Okabe et al. 2015; Zamay et al. 2017), raising protocol of biomarker-driven therapies (Rudman et al. 2020), or genetic information mainly based on single cell mRNA or DNA basis (Li et al. 2022; He et al. 2022; Zhang et al. 2024; Bu et al. 2025) have been reported. They have gained attention as new prognostic biomarkers in a variety of cancers. These methods are included in a cross-sectional approach, such as overall or disease-free survival at certain time points, such as 30 or 90 days, or five or ten years, at the level of the stratified patient group. The hidden problem with cross-sectional methods is that they only show whether a patient is alive at a specific moment in time. They are not able to predict how long an individual will survive. As far as we know, the Cally Index is one of the few binary categorization methods. The Cally Index uses C-reactive protein, albumin, and lymphocytes to predict whether patients with levels above or below the cutoff point are alive or dead (Mizota et al. 2026). However, no longitudinal method is available to predict how long each patient will live. Therefore, it is impossible to predict the long-term outcome of each case.

The purpose of the survival prediction formula, which is the focus of this article, is to help avoid over- or under-treatment and reduce the incidence of treatment-related adverse events. Furthermore, from a medical perspective, the survival time prediction signature can be used to evaluate the advances in general medical treatment. Accurate prediction of postoperative survival time in patients with LUSC may facilitate risk stratification, optimize adjuvant treatment strategies, individualize follow-up schedules, and support clinical decision-making.

Aim

The aim of this study was to create a prediction formula for survival after LUSC surgery using accessible biomarkers. It may provide optimal, personalized medical care and save on costs associated with adverse events that are consistent with post-surgical treatments. Then, we examined the formula's validity.

Methods

Eligibility and Exclusion Criteria for Participants

Patients were included if they were male and diagnosed with LUSC. They also had to undergo radical surgery at either Osaka Medical and Pharmaceutical University Hospital (Hospital A) between April 2007 and December 2015 or Yao Tokushukai General Hospital (Hospital B) between April 2015 and December 2018. Additionally, patients with LUSC who underwent surgery at Hospital A in December 2019 were included as validation test participants for the analysis of the formula developed in Method 2. These time periods were chosen so that the follow-up period would exceed five years after radical surgery. The exclusion criteria are as follows: 1. Females. 2. Males who lost follow-up five years after radical surgery for LUSC. 3. The pathology diagnosis was not squamous cell carcinoma.

Variables Available for Prediction Formula Development

The variables collected from the subject's electronic medical records were as follows: 1, **Demographics**, including sex, age, body mass index (BMI, body weight (kg) divided by height (meters) square, kg / m^2), 2, **Laboratory data** of complete blood counts (CBC) and biochemistry, including white blood cell, total lymphocyte (TLC), neutrophil (Neu), hemoglobin (Hb), platelet (PLT), and albumin (Alb). Blood measurements were taken within 28 days of hospitalization prior to radical LUSC surgery, 3, **Biomarker indices**, that were readily available, were chosen, including inflammation index and surgical stress index - advanced lung cancer inflammation index (ALI) (Bu et al. 2025), Hb, Alb, TLC, PLT (HALP) score (Jafri et al. 2013; Gursoy et al. 2024), Neu-lymphocyte ratio (NLR) (Buonacera et al. 2022), and systemic immune-inflammation index (SII) (Selahattin et al. 2023). The equations for calculating them are shown downwards, 4, **Pathological measures** included pathological TNM stage (pTNM stage) and PDPN immune-histological staining results were evaluated. Tumor stage was defined according to the Union for International Cancer Control (UICC) TNM classification, and histologic type was defined according to the World Health Organization classification.

The calculation formulas for ALI, HALP, NLR, and SII, as used in this method, are shown below:

Advanced Lung Cancer Inflammation Index (ALI) = $\text{BMI} \times \text{Alb} / \text{NLR}$ (Bu et al. 2025)

Neutrophil/Lymphocyte Ratio (NLR) = Neu / TLC (Buonacera et al. 2022)

Hemoglobin Albumin Lymphocyte Platelets (HALP) = $\text{Hb} \times \text{Alb} \times \text{TLC} / \text{PLT}$ (Chen et al. 2015; Gursoy et al. 2024)

Systemic immune-inflammation index (SII) = $\text{PLT} \times \text{NLR}$ (Selahattin et al. 2023).

The abbreviations used here are as follows: Alb: serum albumin concentration, BMI: body mass index, calculated by dividing body weight (in kilograms) by height squared (in meters^2), Neu: neutrophil count, PLT: platelet count, TLC: total lymphocyte count.

Outcome measure

The outcome measure was 5-year survival, defined as alive or dead at the time of 5 years after surgery for LUSC.

Sample size calculation

First, to calculate sample size in the developing prediction formula in phase 1, with a 95% confidence level and 5% margin error, and population proportion and population size, we set at 20% and 100, respectively. From these settings, the sample size was calculated as 72. Second, to calculate the sample size for the external validation test in phase 2, the number of subjects was set at 70, the confidence level was set at 90%, and the margin of error was set at 10%. As a result, the number of subjects required was 35, where 35 was set as the lowest limit of subject number of validation test.

Assessment of inter-institutional heterogeneity

Prior to pooling the datasets from two hospitals, inter-institutional differences were evaluated. Baseline characteristics were compared using the chi-squared test (χ^2) where appropriate, and overall survival was assessed using the Kaplan–Meier method and compared using the log-rank test. Furthermore, a multivariable Cox proportional hazards model including institution as a covariate was used to assess potential center effects. If no clinically meaningful differences were observed, the datasets were combined for subsequent analyses.

Method 1. Development of the Prediction formula -Phase 1

Method 1.1. Comparison of data of survival and death within 5 years of LUSC surgery

All cases were divided into two groups, a 5-year survival group and a death group at the time of 5 years passed after the radical operation. Then, all collected data were compared to discriminate them to draw the biomarker to predict survival cases.

Method 1.2. Development of the Prediction Formula of survival time

Using the variable for predicting survival days after the LUSC radical operation, we developed the prediction formula using Cox proportional hazards regression analysis. Receiver operating characteristic (ROC) curve analysis was also performed to determine the cutoff value of the biomarkers available in the prediction formula.

Method 2. Validation of Model discrimination for the Developed Prediction Formula

Model discrimination was assessed using Harrell's concordance index (C-index). Model calibration was evaluated using calibration plots and the calibration slope. Internal validation was performed using bootstrap resampling with 1,000 repetitions to estimate optimism-corrected performance. External validation was conducted using an independent cohort, in which model discrimination and calibration were also evaluated. Clinical usefulness was assessed using decision curve analysis. This study was reported in accordance with the TRIPOD Statement.

Method 3. Kaplan-Meier analysis for the possibility of Prediction Using PDPN Staining Results

The correlation between PDPN staining and survival rate was investigated. Patients' surgical specimens were analyzed by histological staging and PDPN staining strength, divided into three degrees of strong, weak positive, or negative. The details are described below. A Kaplan-Meier analysis was performed to compare two groups divided by PDPN staining results: strong positive versus the remaining, which included weak positive and negative staining.

PDPN staining method

LUSC surgical specimen was collected from each case at the time of surgery, and fixed in 10% formalin for 24–48 hours, embedded in paraffin, sectioned at 4 μ m, and stained with hematoxylin and eosin (H&E). For histopathological evaluation by PDPN staining, thin surgical sections of formalin-fixed paraffin-embedded tissue blocks were used. The immunohistochemical techniques were performed using anti- PDPN antibody (D2-40, DAKO) to evaluate its expression. Lymphatic endothelium was used as a strong positive control, and the presence of cancer cells stained with anti- PDPN antibody resulted in PDPN positive results. PDPN staining results were classified into three as follows: strong positive (Supplementary Fig. 1), weak positive (Supplementary Fig. 2), and negative (Supplementary Fig. 3). All immunohistochemical data were evaluated by a single pathologist (S.N.) who was blinded to the clinical status of the patients.

Statistical analysis

Prognostic factors independent 9ncluding sex, age, Alb, CRP, Hb, PLT, TLC, Neu, monocyte count, ALI, HALP, NLR, and SII as biomarker indicators were examined. Overall survival rate (OS) was calculated from the date of death from any cause after surgery. Survival was analyzed using a Cox proportional hazards regression model. Variables with $P < 0.05$ in univariate analysis were entered into the multivariate model. Hazard ratios (HRs) and 95% confidence intervals (CIs) were calculated. Proportional hazards assumptions were assessed using Schoenfeld residuals. A prognostic formula was constructed based on the regression coefficients of the final multivariate model. The details of the validation analyses used to validate the predictive formula are described above. $P < 0.05$ was considered a statistically significant difference. All statistics were performed using SPSS version 29 (IBM, Armonk, NY, USA).

Results

77 cases were included in the study. Among them, the result of sample size calculation is 72 as described in the statistics session. Among 77 cases, 58 cases were from Hospital A and 19 cases were from Hospital B. To see female subjects, 6 from Hospital A and 2 from Hospital B, a total of 8 females were excluded. As the remaining 52 and 17 from two hospitals, a total of 69 cases were proceeded to further analysis (Fig. 1). The postoperative follow-up period was set at 60 months. The median age was 73.5 years (IQR: 55, 86 years). In terms of pathological stage classification, IA, IB, IIA, IIB, IIIA, IIB, and IV were 20, 23, 3, 13, 9, 0, and 1 case, respectively. There were 14 cases with lymph node metastasis.

Result of Assessment of inter-institutional heterogeneity

Data from the two institutions were compared to examine the demographics, laboratory data, and outcomes of patients who underwent LUSC curative surgery at the two institutions. No statistically significant differences were found (Table 1). Therefore, it was determined that all data from the two institutions could be combined for further analysis. Therefore, data from two institutions were combined for analysis in the following methods.

Table 1
Comparison of data of patients with LUSC in two hospitals: Hospital A and B.

Measures		Total	Hospital A	Hospital B	p Value
Subject number	Subject number	69	52	17	
Demographics	Age, Years old	73 (70, 78)	72 (70, 77)	75 (72, 80)	0.149
	BMI, kg/m ²	22.0 (20.0, 24.3)	22.3 (20.2, 25.0)	20.8 (19.0, 23.7)	0.124
Laboratory index	Alb, g/dL	4.1 (3.6, 4.3)	4.1 (3.7, 4.3)	3.9 (3.6, 4.2)	0.096
	CRP, mg/dL	0.20 (0.10, 1.18)	0.20 (0.05, 0.97)	0.28 (0.12, 1.48)	0.342
	Hb, g/dL	13.5 (12.1, 14.5)	13.6 (12.3, 14.4)	13.5 (11.3, 14.8)	0.983
	PLT, 10 ⁴ /μL	22.5 (18.2, 27.1)	22.7 (18.8, 28.1)	20.1 (17.1, 24.7)	0.166
	TLC, counts/μL	1667 (1268, 2049)	1606 (1289, 2100)	1825 (1111, 2006)	0.759
	Neutrophile, counts/μL	4108 (2888, 4966)	4217 (2880, 5183)	3784 (2867, 4829)	0.646
	Mono, counts/μL	388 (302, 494)	389 (294, 507)	372 (312, 470)	0.749
Calculated biomarker	ALI	36.02 (26.01, 46.34)	35.65 (26.99, 46.91)	38.30 (24.75, 46.20)	0.759
	HALP	37.87 (27.01, 50.71)	36.63 (26.77, 51.63)	38.97 (27.78, 47.17)	0.989
	NLR	2.30 (1.78, 3.38)	2.30 (1.79, 3.42)	2.35 (1.64, 3.35)	0.754
	PLR	143.89 (111.69, 170.93)	149.10 (111.85, 176.86)	136.76 (97.71, 162.26)	0.373

There were no significant differences in all measured data from the two hospitals.

Abbreviations, Alb: serum albumin, ALI: Advanced Lung Cancer Inflammation Index, BMI: body mass index, CRP: C-reactive protein, HALP: Hemoglobin Alb Lymphocyte Platelets, Hb: hemoglobin, NLR: neutrophile-lymphocyte ratio, PDPN: Podoplanin, PLR: platelet lymphocyte ration, SII: Systemic immune-inflammation index, TLC: total lymphocyte count.

Measures		Total	Hospital A	Hospital B	p Value
Subject number	Subject number	69	52	17	
Demographics	Age, Years old	73 (70, 78)	72 (70, 77)	75 (72, 80)	0.149
	SII, 10 ⁴ /μL	55.4 (37.4, 76.1)	56.5 (38.0, 80.4)	44.4 (31.5, 68.3)	0.392
Pathological findings	p Stage $\geq$ A and I B, n (%)	43 (62)	33 (63)	10 (59)	0.732
	PDPN strong positive, n (%)	13 (19)	9 (17)	4 (24)	0.402
Outcome	5 year survival days	1825 (684, 1825)	1825 (676, 1825)	1324 (708, 1825)	0.356
There were no significant differences in all measured data from the two hospitals.					
Abbreviations, Alb: serum albumin, ALI: Advanced Lung Cancer Inflammation Index, BMI: body mass index, CRP: C-reactive protein, HALP: Hemoglobin Alb Lymphocyte Platelets, Hb: hemoglobin, NLR: neutrophile-lymphocyte ratio, PDPN: Podoplanin, PLR: platelet lymphocyte ration, SII: Systemic immune-inflammation index, TLC: total lymphocyte count.					

Result 1.1. Comparison of Survival and Death Groups in Developing Set

Comparing the survival and death groups, statistically significant differences were observed in subject demographics and blood test, including BMI, Alb, and Hb (Table 2).

Table 2
Comparison of data of survivors and death at the 5-year after the radical operation.

Measures		Total	Survivors	Deaths	p-value
Subject number		69	38	31	
Demographics	Age, Years old	73 (70, 78)	72 (68, 78)	73 (71, 78)	0.246
	BMI, kg/m ²	22.0 (20.0, 24.3)	23.3 (21.0, 24.9)	20.8 (18.8, 23.7)	0.004
Laboratory index	Alb, g/dL	4.1 (3.6, 4.3)	4.1 (3.9, 4.4)	3.8 (3.5, 4.2)	0.014
	CRP, mg/dL	0.20 (0.10, 1.18)	0.15 (0.07, 0.73)	0.40 (0.10, 1.87)	0.075
	Hb, g/dL	13.5 (12.1, 14.5)	14.0 (12.9, 15.0)	12.4 (11.1, 14.2)	0.002
	PLT, 10 ⁴ /μL	22.5 (18.2, 27.1)	22.7 (17.7, 26.8)	22.0 (18.7, 28.2)	0.942
	TLC, counts/μL	1667 (1268, 2049)	1724 (1387, 2225)	1532 (1156, 1985)	0.132
	Neutrophile, counts/μL	4108 (2888, 4966)	4035 (2905, 5054)	4192 (2871, 4943)	0.856
	Mono, counts/μL	388 (302, 494)	351 (288, 492)	417 (347, 504)	0.221
Calculated biomarker	ALI	36.02 (26.01, 46.34)	40.77 (31.07, 57.95)	32.61 (20.58, 39.90)	0.003
	HALP	37.87 (27.01, 50.71)	43.94 (31.48, 55.92)	35.57 (23.50, 40.86)	0.009
	NLR	2.30 (1.78, 3.38)	2.25 (1.53, 3.27)	2.43 (2.08, 3.76)	0.144

When the demographic data and biomarkers of the survivors and deceased were compared, it was found that BMI, Hb, ALI and HALP were significantly lower in the deceased group.

Abbreviations, Alb: serum albumin, ALI: Advanced Lung Cancer Inflammation Index, BMI: body mass index, CRP: C-reactive protein, HALP: Hemoglobin Alb Lymphocyte Platelets, Hb: hemoglobin, NLR: neutrophile-lymphocyte ratio, PDPN: Podoplanin, PLR: platelet lymphocyte ration, SII: Systemic immune-inflammation index, TLC: total lymphocyte count.

Measures		Total	Survivors	Deaths	p-value
Subject number		69	38	31	
Demographics	Age, Years old	73 (70, 78)	72 (68, 78)	73 (71, 78)	0.246
	PLR	143.89 (111.69, 170.93)	125.42 (99.35, 165.72)	152.37 (134.55, 210.53)	0.070
	SII, 10 ⁴ /μL	55.4 (37.4, 76.1)	52.9 (33.4, 72.2)	60.9 (37.8, 92.4)	0.167
Pathological findings	p Stage ≠ AB, n (%)	43 (62)	27 (71)	16 (52)	0.097
	PDPN strong positive, n (%)	41 (59)	24 (63)	17 (55)	0.484
Outcome	Survival days	1826	1826	621	< 0.001
	After operation	(684, 1826)	(1826, 1826)	(331, 1133)	
When the demographic data and biomarkers of the survivors and deceased were compared, it was found that BMI, Hb, ALI and HALP were significantly lower in the deceased group.					
Abbreviations, Alb: serum albumin, ALI: Advanced Lung Cancer Inflammation Index, BMI: body mass index, CRP: C-reactive protein, HALP: Hemoglobin Alb Lymphocyte Platelets, Hb: hemoglobin, NLR: neutrophile-lymphocyte ratio, PDPN: Podoplanin, PLR: platelet lymphocyte ration, SII: Systemic immune-inflammation index, TLC: total lymphocyte count.					

Result 1.2. Cox proportional hazards regression analysis for detecting Variables as confound available in the prediction formula

To identify the available biomarkers to discriminate survival and death within 5 years after surgery, multiple logistic regression analysis for relative hazard ratio (RR), 95% confidence interval (95% CI) and p-value for predicting 5-year survival was performed (Table 3). As the results, ALI and Hb were independent variables for 5-year survival prediction. The results of comparisons of HALP, NLR, and SII had no significant RR.

Table 3

The results of the multiple logistic regression analysis, which detects confounding variables, are available in the prediction formula for survival time.

Variable	OR (95% CI)	p-value
Age	1.003 (0.920–1.093)	0.945
ALI	0.961 (0.928–0.996)	0.030
Hb	0.670 (0.474–0.946)	0.023
PDPN	0.987 (0.319–3.057)	0.982
Of the four variables, only ALI and Hb remained for use in developing a formula to predict survival time in patients with LUSC.		
Abbreviations, ALI: Advanced Lung Cancer Inflammation Index, CI: confidence interval, Hb: hemoglobin, LUSC: lung squamous cell carcinoma, OR: odds ratio.		

Result 1.3. Development of the prediction formula of survival time

In Result 1.2.1, as variables with $P < 0.05$ in univariate analysis were ALI and Hb, both of them were entered into the Cox proportional hazards regression model. As the result, the prediction formula of 5-year survival days after surgery for patients with LUSC was drawn as follows:

$$\text{Survival time (days)} = [6.9 \times \text{ALI}] + [93.4 \times \text{Hb}] - 198.4.$$

The time-dependent ROC curve analysis was performed to determine the cutoff value of ALI for predicting 5-year survival. As a result, the cutoff value of ALI was 48.04 (AUC = 0.690, $p = 0.003$). These statistical results of the ROC curve analysis are interpreted that this prediction formula is acceptable for clinical availability.

The Kaplan–Meier survival curves stratified by predicted survival time calculated using the ALI–Hb prediction model. Patients were categorized into high-(< 1,000), and low- (< 1,000) risk groups according to predicted survival duration. The model demonstrated clear separation of survival outcomes among the two groups (Fig. 2).

Result 2. Assessment of Model discrimination for the Developed Prediction Formula

Discrimination and Calibration for Internal Validation

The prediction model demonstrated acceptable discrimination in the development cohort, with a Harrell's C-index of 0.70 (95% CI, 0.66–0.74). Calibration plots showed good agreement between predicted and observed survival probabilities across risk groups. The calibration slope was 0.95,

indicating minimal overfitting. Bootstrap internal validation using 1,000 resamples yielded an optimism-corrected C-index of 0.68, suggesting stable model performance with limited optimism. The calibration slope after internal validation remained close to unity, supporting the robustness of the model. In the external validation cohort, the model maintained acceptable discrimination, with a C-index of 0.65 (95% CI, 0.61–0.69). Calibration assessment demonstrated good agreement between predicted and observed outcomes, with no substantial deviation across risk strata. Patients were stratified into high- and low-risk groups based on the predefined cutoff value. Kaplan–Meier survival analysis revealed a significant difference in survival between the two groups (log-rank test, $P = 0.0013$, Fig. 4), indicating the clinical usefulness of the model for risk classification. Overall, these findings indicate that the model achieved moderate but consistent predictive performance across both internal and external validation settings.

Summary of key findings and Clinical implications

Taken all obtained results together, these results indicate that a prognostic model incorporating ALI and Hb can stratify patients with LUSC into distinct risk categories with different survival outcomes. Using ALI and Hb, the prediction formula for survival days after radical surgery has been developed: Survival time (days) = $[6.9 \times \text{ALI}] + [93.4 \times \text{Hb}] - 198.4$. The prediction formula was proven to be favorable through verification of internal and external validity.

Result 3. Results of Investigation of Correlation Between PDPN Staining Results and Survival

Staining surgical LUSC specimens with the PDPN antibody revealed 41 positive cases (13 strong and 28 weak) and 28 negative cases. In the next step, all cases were divided into two subgroups, one with strong PDPN and the other with weak and negative PDPN. Comparing all data including survival between two groups, statistically significant differences were found in age (p-value 0.026) and SII (p-value 0.044) (Table 4). Then, a Kaplan-Meier analysis was performed to compare the two groups. The results of the analysis revealed a trend toward longer survival in the strong positive group. This result suggests that there was a survival trend among PDPN strong positive patients that was better than that of the others. However, there was no significant difference in survival rates between the two groups.

Table 4

A comparison of the outcome measures between the results of PDPN staining, with a strong positive result compared to a weak positive or negative result.

Subject number	PDPN staining	Total	Strong positive	Weak positive or negative	p Value
		69	13	56	
Demographics	Age, Years old	73 (70, 78)	75 (73, 83)	72 (69, 78)	0.026
	BMI, kg/m ²	22.0 (20.0, 24.3)	22.4 (18.7, 25.3)	22.0 (20.0, 24.1)	0.921
Biomarker	Hb, g/dL	13.5 (12.1, 14.5)	13.5 (12.6, 15.1)	13.6 (11.8, 14.5)	0.612
Calculating biomarker	ALI	36.02 (26.01, 46.34)	46.27 (33.53, 69.73)	35.39 (24.07, 45.14)	0.053
	HALP	37.87 (27.01, 50.71)	37.87 (30.89, 79.32)	37.52 (24.70, 50.23)	0.349
	SII, 10 ⁴ /μL	55.4 (37.4, 76.1)	39.3 (29.1, 59.2)	59.1 (38.4, 80.4)	0.044
Pathological stage	p Stage IA and IB, n (%)	43 (62)	10 (77)	33 (59)	0.189
Outcome measures	5 years surviving time, day	1826 (684, 1826)	1826 (1422, 1826)	1826 (554, 1826)	0.091
	Overall survival rate, n (%)	38 (55)	9 (69)	29 (52)	0.255
The differences observed in the PDPN staining results were related to age and SII. However, no differences were observed in Hb or ALI.					
Abbreviations, ALI: Advanced Lung Cancer Inflammation Index, BMI: body mass index, HALP: Hemoglobin Alb Lymphocyte Platelets, PDPN: podoplanin, SII: Systemic immune-inflammation index.					

Discussion

Principal findings

In this retrospective study, we developed a prognostic model for overall survival in patients with LUS. Among the evaluated clinical and laboratory variables, ALI and Hb were identified as independent prognostic factors, and the predictive formula of the survival time (days), 'Survival time (days) = [6.9 × ALI] + [93.4 × Hb] - 198.4', enabled stratification of patients according to survival outcomes. Previous

studies have reported an association between ALI and prognosis in lung cancer patients. Our findings support the biological and clinical plausibility of the proposed model.

Interpretation of findings

The observed C-index values indicate acceptable discrimination, which is comparable to previously reported clinical prediction models in oncology. Importantly, the calibration slope remained close to 1.0, suggesting that the model predictions were not substantially overfitted and were well aligned with observed outcomes. Although the discrimination was not exceptionally high, this level of performance is not unexpected in survival prediction models based on routinely available clinical variables. Complex biological processes underlying cancer progression often limit the achievable predictive accuracy of models derived from standard clinical data alone.

Clinical implications

From a clinical perspective, the model may be useful for risk stratification and individualized postoperative management. The ability to classify patients into distinct risk groups, as demonstrated by the Kaplan–Meier analysis, could support clinical decision-making, including follow-up intensity and consideration of adjuvant therapies.

Literature Review of ALI-Related Studies

The Advanced Lung Cancer Inflammation Index (ALI) was originally developed as a prognostic indicator for patients with stage IV non-small cell lung cancer (NSCLC). Since its introduction, ALI has been increasingly investigated as a prognostic biomarker in a variety of malignancies and chronic inflammatory diseases.

As of March 1, 2026, a literature search of PubMed using “ALI” as a keyword identified 261 articles. After screening, 165 articles were excluded because they were not related to the Advanced Lung Cancer Inflammation Index. The remaining 96 studies evaluated ALI as a prognostic indicator across different disease categories (Table 5), including lung cancer, non-lung cancers, lifestyle-related diseases, and other inflammatory conditions. Among the lung cancer studies, most focused on NSCLC, whereas no studies specifically examined lung squamous cell carcinoma (LUSC). Importantly, no previous studies have reported a survival prediction formula incorporating ALI or related inflammatory indices. Beyond lung cancer, ALI has been widely investigated in non-lung malignancies, particularly gastrointestinal cancers. In addition, several studies have demonstrated its prognostic relevance in lifestyle-related diseases such as heart failure, asthma, acute coronary syndrome, and myocardial infarction. These observations suggest that ALI reflects systemic biological processes that are broadly associated with disease prognosis.

Table 5

The 96 ALI-related articles were grouped into four categories of disease: lung cancer, non-lung cancer, life-related diseases and other diseases.

Four Categories	Diseases	Total number of articles	Sub-total number of articles	Supplementary references
Lung cancer		29		
	NSCLC		16	6, 15, 21, 28, 37, 41, 46, 52, 55, 62, 73, 75, 78, 84, 87, 94.
	SCLC		4	32, 61, 68, 93.
	Lung cancer		7	2, 7, 34, 49, 82, 83, 90.
	Lung adenocarcinoma		2	29, 69.
Non-lung cancer		34		
	Gastrointestinal tract		14	3, 8, 14, 16, 22, 33, 42, 44, 56, 63, 65, 74, 79, 95.
	HCC		6	9, 24, 30, 39, 50, 76.
	Head Neck		2	35, 59.
	MM		2	43, 57.
	Urinary tract		2	45, 92.
	Malignant lymphoma		2	86, 96.
	Pancreas		2	54, 71.
	Cholangiocarcinoma		1	27
	Skin		1	26
	Neuroblastoma		1	25

ALI-related 96 articles were grouped into four categories: lung cancer-related articles, non-lung cancer-related, lifestyle-related diseases, and other diseases of article number of 29, 34, 9, and 34, respectively. Interestingly no ALI-related articles were found for LUSC. However, 34 articles related to ALI were found in non-lung cancers, with the largest number (14 articles) relating to gastrointestinal cancer. The other subject of ALI-related articles to use ALI to predict the prognosis was lifestyle-related diseases, including heart failure, asthma, acute coronary syndrome and myocardial infarction.

Abbreviations, ACS: acute coronary syndrome, ALI: advanced lung cancer inflammation index, CAP: community-acquired pneumonia, CKD: chronic kidney disease, DM: diabetes mellitus, HCC: hepatic cell carcinoma, HT: hypertension, MASLD: Metabolic Dysfunction Associated Steatotic Liver Disease, MI: myocardial infarction, MM: malignant melanoma, NSCLC: non-small cell lung cancer, RA: rheumatic arthritis, SCLC: small cell lung cancer.

Four Categories	Diseases	Total number of articles	Sub-total number of articles	Supplementary references
Lung cancer		29		
	NSCLC		16	6, 15, 21, 28, 37, 41, 46, 52, 55, 62, 73, 75, 78, 84, 87, 94.
	Unknown		1	47
Life-related diseases		9		
	HT		2	5, 11.
	DM		2	1, 36.
	CKD		3	10, 40, 64.
	MASLD		2	72, 91.
Others		24		
	Heart failure		5	12, 13, 18, 19, 23.
	Asthma		4	31, 38, 51, 85.
	ACS		3	20, 67, 89.
	MI		3	60, 66, 88.
	Others		9	4, 58 (stroke) 17, 48 (RA) 53 (pulmonary fibrosis) 70 (CAP) 77 (pneumonia) 80 (chronic pain)

ALI-related 96 articles were grouped into four categories: lung cancer-related articles, non-lung cancer-related, lifestyle-related diseases, and other diseases of article number of 29, 34, 9, and 34, respectively. Interestingly no ALI-related articles were found for LUSC. However, 34 articles related to ALI were found in non-lung cancers, with the largest number (14 articles) relating to gastrointestinal cancer. The other subject of ALI-related articles to use ALI to predict the prognosis was lifestyle-related diseases, including heart failure, asthma, acute coronary syndrome and myocardial infarction.

Abbreviations, ACS: acute coronary syndrome, ALI: advanced lung cancer inflammation index, CAP: community-acquired pneumonia, CKD: chronic kidney disease, DM: diabetes mellitus, HCC: hepatic cell carcinoma, HT: hypertension, MASLD: Metabolic Dysfunction Associated Steatotic Liver Disease, MI: myocardial infarction, MM: malignant melanoma, NSCLC: non-small cell lung cancer, RA: rheumatic arthritis, SCLC: small cell lung cancer.

Four Categories	Diseases	Total number of articles	Sub-total number of articles	Supplementary references
Lung cancer		29		
	NSCLC		16	6, 15, 21, 28, 37, 41, 46, 52, 55, 62, 73, 75, 78, 84, 87, 94.
				81 (gallstone)
Total number of articles		96		
ALI-related 96 articles were grouped into four categories: lung cancer-related articles, non-lung cancer-related, lifestyle-related diseases, and other diseases of article number of 29, 34, 9, and 34, respectively. Interestingly no ALI-related articles were found for LUSC. However, 34 articles related to ALI were found in non-lung cancers, with the largest number (14 articles) relating to gastrointestinal cancer. The other subject of ALI-related articles to use ALI to predict the prognosis was lifestyle-related diseases, including heart failure, asthma, acute coronary syndrome and myocardial infarction.				
Abbreviations, ACS: acute coronary syndrome, ALI: advanced lung cancer inflammation index, CAP: community-acquired pneumonia, CKD: chronic kidney disease, DM: diabetes mellitus, HCC: hepatic cell carcinoma, HT: hypertension, MASLD: Metabolic Dysfunction Associated Steatotic Liver Disease, MI: myocardial infarction, MM: malignant melanoma, NSCLC: non-small cell lung cancer, RA: rheumatic arthritis, SCLC: small cell lung cancer.				

The prognostic value of ALI may be explained by its integration of body mass index (BMI), serum albumin levels, and lymphocyte count, which collectively reflect nutritional status and systemic inflammation. Cancer progression is strongly influenced by both malnutrition and chronic inflammation, and ALI may provide a simple composite indicator that captures these host-related factors. Despite the growing number of ALI-related studies, the mechanisms underlying its association with cancer outcomes remain incompletely understood. Further research is therefore required to clarify the biological basis of ALI and to determine its role in prognostic modeling for lung cancer. To our knowledge, this is the first study to develop and validate a survival prediction formula incorporating ALI in patients with LUSC.

Lack of Clinical Implications Regarding the Predictive Value of PDPN

Podoplanin (PDPN) is a mucin-type transmembrane sialoglycoprotein that is widely used as a histopathological marker to distinguish lymphatic vessels from blood vessels because it is expressed in lymphatic endothelial cells (Schoppmann et al., 2001; Breiteneder-Geleff et al., 1999). In several studies, PDPN-positive tumors have been associated with an unfavorable prognosis, possibly due to the formation of an immunosuppressive tumor microenvironment (Sasaki et al., 2018; Suzuki et al., 2021). However, other studies have reported the opposite association, suggesting that the prognostic significance of PDPN remains controversial (Hu et al., 2020).

In the present study, we aimed to clarify whether PDPN expression is associated with survival outcomes in our cohort. However, contrary to our expectations, PDPN expression was not significantly associated with survival. These findings suggest that the clinical utility of PDPN as a prognostic biomarker in lung squamous cell carcinoma (LUSC) may be limited. Further studies with larger sample sizes are required to clarify the prognostic significance of PDPN expression in LUSC.

Possible Explanation for the Discrepancy with Previous Studies

The discrepancy between our findings and those of previous studies may be explained by several factors. First, differences in patient populations and clinical characteristics may influence the prognostic impact of PDPN expression. Many previous studies included heterogeneous groups of lung cancer or multiple histological subtypes, whereas the present study specifically focused on lung squamous cell carcinoma (LUSC). Because tumor biology differs among histological subtypes, the prognostic significance of PDPN may vary accordingly. Second, methodological differences in the assessment of PDPN expression may also contribute to the inconsistent findings. Variations in immunohistochemical staining protocols, antibody selection, and cut-off values used to define PDPN positivity may affect the reported associations with survival outcomes. Third, the relatively limited sample size in the present study may have reduced the statistical power to detect a significant association between PDPN expression and survival. Larger multicenter studies are therefore required to further clarify the prognostic value of PDPN in patients with LUSC.

Strength

The strength of this study is, to our knowledge, the first report to predict survival days in patients with LUSC using ALI and Hb. The predictive indicators currently used for overall or disease-free survival are survival rates at various time points, such as 30 or 90 days, five or ten years, and so on, within the patient group. This index is available as a cross-sectional indicator dealing with various groups of diseases, as discussed in the previous session. The problem with this indicator is that it only shows whether a patient is alive at a given moment, and it cannot predict how many days a patient will survive. In contrast, the survival day prediction formula obtained in the current study is a longitudinal indicator of individual survival days. The strength of the developing formula may contribute to the five areas outlined in the next section as being of clinical importance to the strategy for medical care of lung cancer.

The five contributors are hereby offered the capability to predict the survival time of patients with LUSC after surgery, as follows: **1. Implementation of precision medicine:** This helps to determine the intensity of postoperative adjuvant therapy. **2. Determining the appropriateness of postoperative adjuvant therapy:** Using a prediction formula to stratify patients into low-, medium-, and high-risk groups (mainly observation, standard adjuvant chemotherapy, and actively considering immunotherapy and clinical trials, respectively) can reduce the incidence of adverse events due to under- or overtreatment. **3. Optimizing follow-up strategies:** For patients with a short predicted survivors, close imaging follow-up

and an emphasis on the early detection of recurrence can be implemented. For those with a long predicted survivors, unnecessary CT exposure and medical costs can be reduced to streamline the allocation of medical resources. **4. Providing patient explanations and decision-making support strategies:** Rather than providing population data, such as a '5-year survival rate, %', specific figures can be presented, such as a median and a 95% confidence interval of survival days. This makes advance care planning (ACP), postoperative treatment selection and rehabilitation plans more concrete. **5. Clinical research stratification tool:** This prediction formula enables risk stratification in clinical trials, real-world data analysis and comparative research with AI models. These five factors could transform lung cancer care strategies by shifting towards a more patient-centered and tailored approach.

Limitations

The limitations of this study must be raised. First, the effectiveness of PDPN as a prognostic predictor could not be clarified. The reasons for this may be related to lack of subjects and retrospective research style. In future studies, it will be necessary to increase the number of subjects and to conduct a prospective study with not only early staged but also advanced staged subjects. The mechanism behind why ALI and Hb are useful in the prediction formula is still unclear. To clarify this, it will be necessary to scientifically clarify the significance of the two indices included in ALI and Hb in terms of LUSC survival time. Second, sample size must be small. This was a two-center retrospective study with a relatively small sample size of 69 patients. Because of the limited number of events, the number of variables included in the multivariable model had to be restricted, and the presence of residual confounding cannot be excluded. Third, treatment strategies may not have been uniform throughout the study period, particularly given advances in systemic therapies, including chemotherapy and immunotherapy. Although initial treatment modality was adjusted for in the analysis, the detailed effects of different treatment regimens could not be fully evaluated. Fourth, the retrospective observational design precludes any causal inference. The model is intended solely for prognostic estimation and does not imply treatment efficacy.

Conclusion

A prognostic model for overall survival in patients with lung squamous cell carcinoma was developed using two routinely available clinical variables, and its validity was proven through internal and external verification. The equation is as follows: survival time (days) = $(6.9 \times \text{ALI}) + (93.4 \times \text{Hb}) - 198.4$. This might provide optimal, personalized medical care and save on costs associated with adverse events that are consistent with post-surgical treatments.

Declarations

Author Contributions declaration

Y.I., A.T. and T.A. formulated the original idea; Yoshio Ichihashi completed the database search, data extraction, and analysis; S.N., K.U., H.H., K.O. T.A. and Y.H. drafted the manuscript; K.S., N.H. and T.K. reviewed and revised the manuscript for important intellectual content; and all authors provided final approval of the version to be submitted. All authors have read and agreed to the published version of the manuscript.

Funding Declaration

Not applicable.

Competing interests

The authors declare no competing interests.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analyzed during this study are included in this article. The datasets generated during and analyzed during the current study are available from the corresponding author on reasonable request.

Ethics declarations

This study was approved by the Institutional Review Board of Osaka Medical and Pharmaceutical University (approval number: 2021-185-3 dated on March 20, 2025). The requirement for obtaining written informed consent from individual participants was waived by the Institutional Review Board because of the retrospective nature of the study. Instead, an opt-out approach was employed, whereby study information was made available on the hospital website and through institutional notices, allowing participants the opportunity to decline participation. No participants opted out of the study. All procedures were performed in accordance with the ethical standards of the institutional research committee and with the Declaration of Helsinki and its later amendments.

The datasets used and/or analyzed during the current study.

The datasets generated and/or analyzed during the current study are not publicly available because they contain potentially identifiable patient information but are available from the corresponding author on reasonable request and with permission of the Institutional Review Board.

References

1. Breiteneder-Gelef, S. et al. Podoplanin—a specific marker for lymphatic endothelium expressed in angiosarcoma. *Verh Dtsch. Ges Pathol.* **83**, 270–275 (1999).
2. Bu, Y. et al. MSR1 in lung squamous cell carcinoma: Prognostic and immunological values in pan- and single-cell analyses and a cohort study. *Int. Immunopharmacol.* **145**, 113811. 10.1016/j.intimp.2024.113811 (2025).
3. Buonacera, A., Stancanelli, B., Colaci, M. & Malatino, L. Neutrophil to Lymphocyte Ratio: An Emerging Marker of the Relationships between the Immune System and Diseases. *Int. J. Mol. Sci.* **23** (7), 3636. 10.3390/ijms23073636 (2022).
4. Chen, J. W. & Dhahbi, J. Lung adenocarcinoma and lung squamous cell carcinoma cancer classification, biomarker identification, and gene expression analysis using overlapping feature selection. *Sci. Rep.* **11** (1), 13323. 10.1038/s41598-021-92725-8.methods (2021).
5. Chen, X. L. et al. Prognostic significance of the combination of preoperative hemoglobin, albumin, lymphocyte and platelet in patients with gastric carcinoma: a retrospective cohort study. *Oncotarget* **6** (38), 41370–41382. 10.18632/oncotarget.5629 (2015).
6. Gursoy, V. et al. HALP scores as a novel prognostic factor for patients with myelodysplastic syndromes. *Sci. Rep.* **14** (1), 13843. 10.1038/s41598-024-64166-6 (2024).
7. He, B. et al. Switched alternative splicing events as attractive features in lung squamous cell carcinoma. *Cancer Cell. Int.* **22** (1), 5. 10.1186/s12935-021-02429-2 (2022).
8. Hu, L. et al. Podoplanin is a useful prognostic marker and indicates better differentiation in lung squamous cell cancer patients? A systematic review and meta-analysis. *BMC Cancer.* **20** (1), 424. 10.1186/s12885-020-06936-9 (2020).
9. Jafri, S. H., Shi, R. & Mills, G. Advance lung cancer inflammation index (ALI) at diagnosis is a prognostic marker in patients with metastatic non-small cell lung cancer (NSCLC): a retrospective review. *BMC Cancer.* **13**, 158 (2013).
10. Li, Q. et al. Molecular profiling of human non-small cell lung cancer by single-cell RNA-seq. *Genome Med.* **14** (1), 87. 10.1186/s13073-022-01089-9 (2022).
11. Lu, T. et al. Trends in the incidence, treatment, and survival of patients with lung cancer in the last four decades. *Cancer Manag Res.* **11**, 943–953. 10.2147/CMAR.S187317 (2019).
12. Miller, H. A., van Berkel, V. H. & Frieboes, H. B. Lung cancer survival prediction and biomarker identification with an ensemble machine learning analysis of tumor core biopsy metabolomic data. *Metabolomics* **57** (2022) (2022). 10.1007/s11306-022-01918-3
13. Mizota, K. et al. The Prognostic Impact of C-Reactive Protein-Albumin-Lymphocyte Index (Cally Index) in Patients with Surgically Resected Non-Small-Cell Lung Cancer. *Ann. Surg. Oncol.* **3** (2), 987–995. 10.1245/s10434-025-18661-3 (2026).
14. Razdan, S. et al. Predictive morbidity and mortality after radical cystectomy using calculators: A comprehensive review of the literature. *Urol. Oncol.* **39** (2), 109–120. 10.1016/j.urolonc.2020.09.032 (2021).

15. Redman, M. W. et al. Biomarker-driven therapies for previously treated squamous non-small-cell lung cancer (Lung-MAP SWOG S1400): a biomarker-driven master protocol. *Lancet Oncol.* **21** (12), 1589–1601. 10.1016/S1470-2045(20)30475-7 (2020).
16. Sakai, T. et al. Link between tumor-promoting fibrous microenvironment and an immunosuppressive microenvironment in stage I lung adenocarcinoma. *Lung Cancer.* **126**, 64–67 (2018).
17. Schoppmann, S. F., Birner, P., Studer, P. & Breiteneder-Geleff, S. Lymphatic microvessel density and lymphovascular invasion assessed by anti-podoplanin immunostaining in human breast cancer. *Anticancer Res.* **21** (4A), 2351–2355 (2001).
18. Selahattin Vural, S., Ali Muhtaroglu, A. & Mert Güngör, M. Systemic immune-inflammation index: A new marker in differentiation of different thyroid diseases. *Med. (Baltim).* **102** (31), e34596. 10.1097/MD.00000000000034596 (2023).
19. Sung, H. et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
20. Suzuki, J. et al. Relationship between podoplanin-expressing cancer-associated fibroblasts and the immune microenvironment of early lung squamous cell carcinoma. *Lung Cancer.* **153**, 1–10 (2021).
21. Wang, X. et al. Establishment and clinical application of the nomogram related to risk or prognosis of hepatocellular carcinoma: A review. *J. Hepatocell Carcinoma.* **10**, 1389–1398. 10.2147/JHC.S417123 (2023).
22. Yu, H., Pang, Z., Li, G. & Gu, T. Bioinformatics analysis of differentially expressed miRNAs in non-small cell lung cancer. *J. Clin. Lab. Anal.* **35**(2), e23588. (2021).
23. Zmay, T. N. et al. Current and prospective protein biomarkers of lung cancer. *Cancers (Basel).* **9** (11), 155. 10.3390/cancers9110155 (2017).
24. Zhang, S., Liu, Y., Liu, K., Hu, X. & Gu, X. A review of current developments in RNA modifications in lung cancer. *Cancer Cell. Int.* **24** (1), 347. 10.1186/s12935-024-03528-6 (2024).
25. Zhou, J., Xu, Y., Liu, J., Yu, J. & Chen, D. Global burden of lung cancer in 2022 and projections to 2050: Incidence and mortality estimates from GLOBOCAN. *Cancer Epidemiol.* **93**, 102693 (2024). doi: 10.1016/j.canep.2024.102693.

Figures

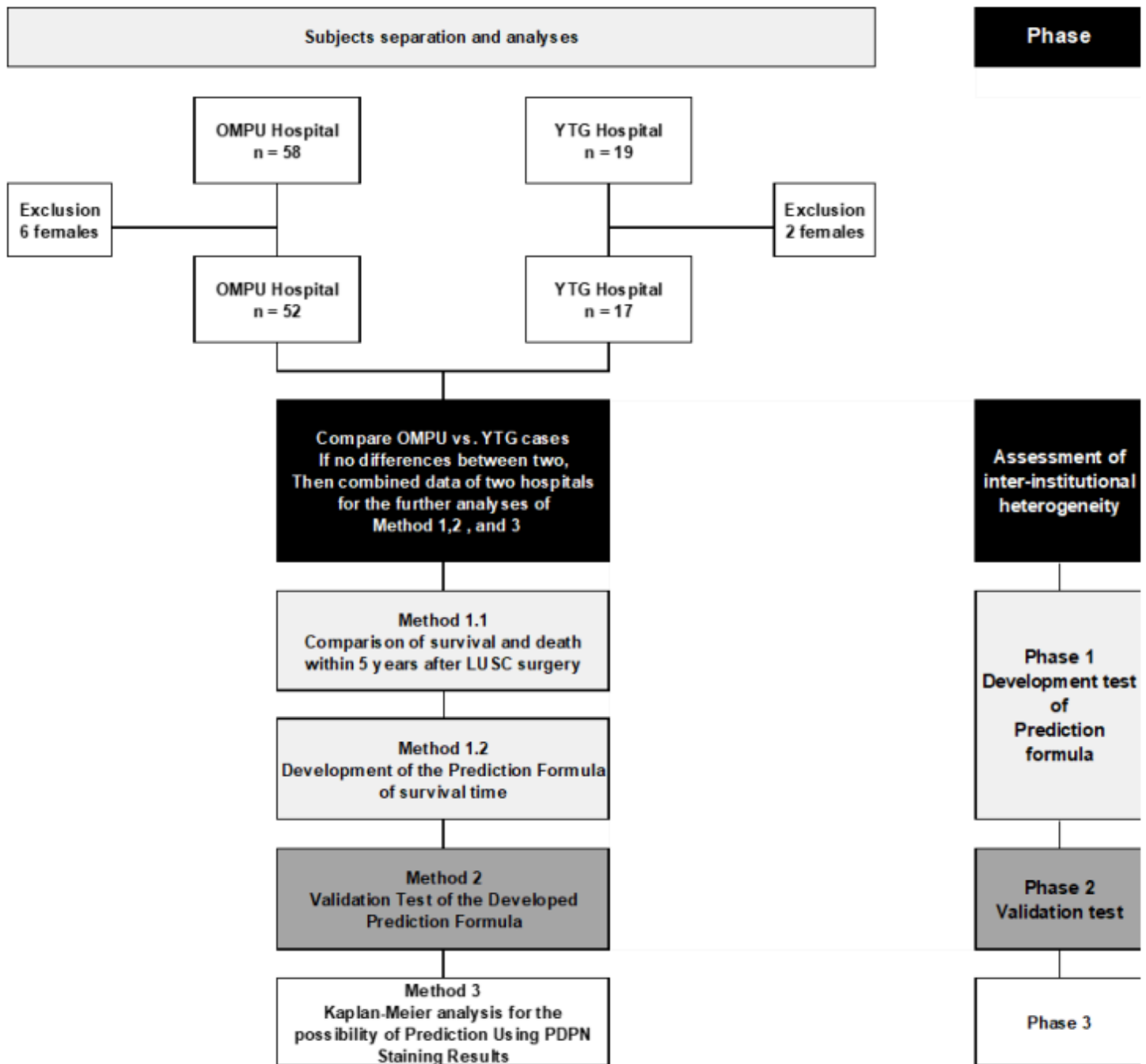


Figure 1

The timeline of the current study. The upper and lower row represents the two hospitals consisting of the study and two phases, respectively.

The data was collected from Osaka Pharmaceutical Medical (OMPU) Hospital (Hospital A) period between 2007 and 2015, in Phase 1 and 2019 in Phase 2. The data was collected also from Yao Tokushukai General (YTG) Hospital (Hospital B) in period between 2015 and 2018 in Phase 1.

Abbreviation, PDPN: podoplanin.

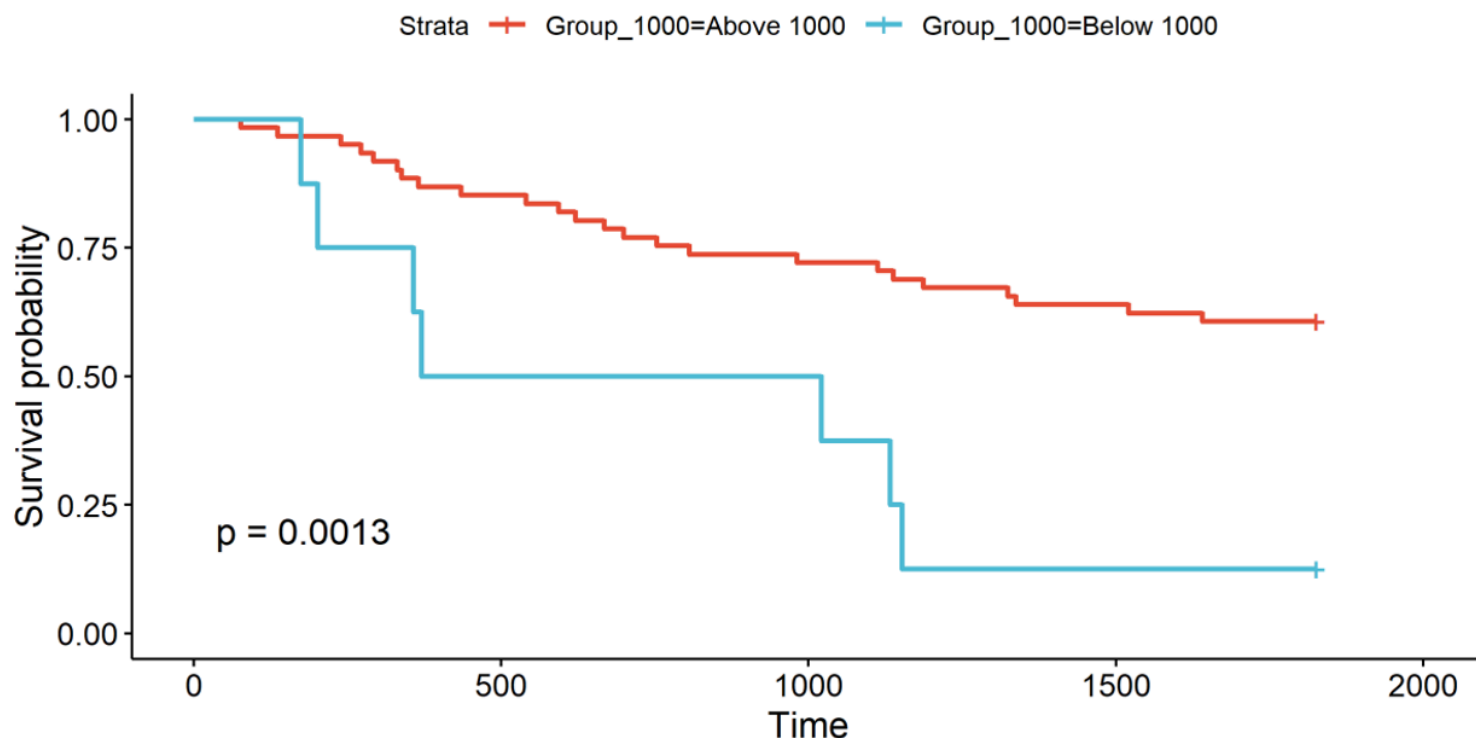


Figure 2

The Kaplan–Meier survival curves stratified by predicted survival time calculated using the ALI–Hb prediction model.

The Kaplan–Meier survival curves stratified by predicted survival time calculated using the ALI–Hb prediction model. Patients were categorized into high- (>1,000), and low- (< 1,000) risk groups according to predicted survival duration. The model demonstrated clear separation of survival outcomes among the two groups.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [ALI261articlesresults.xlsx](#)
- [Alldata.xlsx](#)
- [supplementarytablesandfiles.docx](#)