Am J Cancer Res 2023;13(9):4208-4221
www.ajcr.us /ISSN:2156-6976/ajcr0150771

Original Article

Generation and validation of a predictive model for estimating survival among patients with EGFR-mutant non-small cell lung cancer

Chien-Yu Lin[1*], Yun-Tse Chou[1*], Po-Lan Su[1], Chien-Chung Lin[1,2,3], John Wen-Cheng Chang[4], Chen-Yang Huang[4], Yueh-Fu Fang[5], Ching-Fu Chang[4], Chih-Hsi Scott Kuo[5], Ping-Chih Hsu[5], Cheng-Ta Yang[5], Chiao-En Wu[4]

[1]Department of Internal Medicine, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan 704, Taiwan; [2]Institute of Clinical Medicine, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan 704, Taiwan; [3]Department of Biochemistry and Molecular Biology, College of Medicine, National Cheng Kung University, Tainan 704, Taiwan; [4]Division of Hematology-Oncology, Department of Internal Medicine, Linkou Chang Gung Memorial Hospital, College of Medicine, Chang Gung University, Taoyuan 333, Taiwan; [5]Division of Thoracic Oncology, Department of Thoracic Medicine, Linkou Chang Gung Memorial Hospital, College of Medicine, Chang Gung University, Taoyuan 333, Taiwan. [*]Equal contributors.

Received April 17, 2023; Accepted August 9, 2023; Epub September 15, 2023; Published September 30, 2023

Abstract: Although epidermal growth factor receptor tyrosine kinase inhibitors (EGFR-TKIs) have become the standard therapy for patients with EGFR-mutant non-small cell lung cancer (NSCLC), treatment outcomes vary significantly. Previous studies have indicated that concurrent mutations may compromise the effectiveness of first-line EGFR-TKIs. However, given the high cost of next-generation sequencing, this information is often inaccessible in routine clinical practice. A prediction model based on pre-treatment clinical characteristics may thus offer a more practical solution. This study established a nomogram based on pretreatment clinical characteristics to stratify patients according to optimal treatment strategies. We retrospectively reviewed 761 patients with EGFR-mutant NSCLC who received first- or second-generation EGFR-TKIs at a tertiary referral center between 2010 and 2019. The pretreatment clinical characteristics and progression-free survival data were collected. Using COX proportional hazard regression analysis, we constructed a nomogram based on seven clinically significant prognostic factors: sex, Eastern Cooperative Oncology Group performance status, histology subtype, mutation subtype, stage, and metastasis to the liver and brain. Our nomogram could stratify patients into three groups with different risks for disease progression and was validated in a patient cohort from other hospitals. This risk stratification can provide additional information for determining the optimal first-line treatment strategy for patients with EGFR-mutant NSCLC.

Keywords: Non-small cell lung cancer, epidermal growth factor receptor, tyrosine kinase inhibitors, nomogram

Introduction

For patients with advanced non-small cell lung cancer (NSCLC) harboring epidermal growth factor receptor (EGFR) mutation [1], EGFR tyrosine kinase inhibitors (EGFR-TKIs) remain the mainstay treatment, according to previous phase III studies [2-7]. A subsequent phase 3 FLAURA study further demonstrated the clinical benefits of the third-generation EGFR-TKI osimertinib in both progression-free survival (PFS) and overall survival (OS) compared to first-generation EGFR-TKIs [8, 9]. However, the benefit of OS was not present in every subgroup, especially in the subgroups of patients with Asian ethnicity and exon 21 L858R substitution [9]. In the Giotag study, sequential afatinib and osimertinib provided a median PFS of 27.7 months and a median OS of 37.6 months [10]. The OS was even longer in patients of Asian ethnicity (44.8 months) [10]. These data imply the importance of predicting acquired resistance to T790M point mutations and subsequent osimertinib use. According to the

ARISE study, a longer duration of first-line therapy with first- or second-generation EGFR-TKIs could predict a higher occurrence rate of the acquired T790M mutation during disease progression [11]. Thus, developing a strategy that can prolong or predict the first-line treatment efficacy is important.

Previous studies have demonstrated that the presence of co-occurring mutations, including TP53 mutations, Wnt/β-catenin alterations, and CDK4/6 alterations, might interfere with the treatment efficacy of first-line EGFR-TKIs [12], necessitating more aggressive treatment. Although next-generation sequencing can provide clues for treatment selection, its high cost makes it difficult to apply in real-world settings. A prediction model based on pretreatment clinical characteristics may be more clinically useful.

Recently, Chang et al. identified 11 independent prognostic factors and established a nomogram to stratify EGFR-mutant NSCLC patients into different risk groups [13]. However, this study had only internal validation by bootstrapping and lacked external validation. Our study aimed to establish an externally validated nomogram for predicting PFS.

Materials and methods

Patients and data collection

Patients diagnosed with advanced-stage EGFR-mutant NSCLC between January 2010 and December 2019 were screened retrospectively. Patients who received first-line treatment with EGFR-TKIs in a tertiary referral center, the National Cheng Kung University Hospital (NCKUH), were enrolled as the derivation cohort. Patients with other active malignancies were excluded. Baseline characteristics, including age, sex, Eastern Cooperative Oncology Group (ECOG) performance status (PS) score, smoking history, histology subtype, stage, distant metastasis status, EGFR mutation subtype, and treatment selection of EGFR-TKIs, were recorded in a standardized data collection form.

To examine the generalizability of the nomogram, external validation using an independent patient cohort was also performed. Patients with advanced-stage EGFR-mutant NSCLC who received first-line EGFR-TKIs from January 2016 to December 2017 at Chang Gung Memorial Hospital (CGMH; Linkou, Chiayi, Kaohsiung, and Keelung branches) were retrospectively reviewed and enrolled as the validation cohort. All enrolled patients had sufficient data to score all variables in the established nomograms. The present study was reviewed and approved by the Review Board and Ethics Committee of National Cheng Kung University Hospital (IRB approval number: B-ER-109-043) and Chang Gung Memorial Hospital (IRB approval number: 201901395B0C501). All data were anonymized according to approved guidelines and the Declaration of Helsinki.

Treatment response

After the initiation of EGFR-TKIs, all patients underwent regular examinations with computed tomography of the chest, whereas brain magnetic resonance imaging and bone scan were arranged based on clinical symptoms. PFS, defined as the time from the initiation of EGFR-TKI therapy to the radiological evidence of disease progression, and OS, defined as the time from the initiation of EGFR-TKI therapy to death, were assessed using the Kaplan-Meier curve and compared using the log-rank test.

Statistical analysis

Continuous variables were compared using the analysis of variance, Kruskal-Wallis test, and independent sample t-test. Categorical variables were compared using Pearson's chi-square test or Fisher's exact test. First, we performed univariate analysis to identify potential prognostic factors of PFS and OS according to various pretreatment clinical characteristics. Clinically significant prognostic factors were then incorporated into multivariate Cox proportional hazard regression analysis to assess the multivariable relationships between predictors and 6-, 9-, and 12-month PFS and OS rates.

Nomogram creation and statistical software

Based on the significant variables defined by the multivariate COX regression analysis, nomograms were constructed using the R software (R version 4.0.5, R Core Team, 2021, R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/) with rms package. The validation of model perfor-

Table 1. Baseline characteristics of the derivation and validation cohorts

Characteristics	Dataset		*p*-Value
	Derivation/NCKUH (N = 761)	External validation/CGMH (N = 751)	
Age (years), mean ± SD	65.79 ± 11.57	68.06 ± 11.79	< 0.001
≤ 65	371 (48.8%)	313 (41.7%)	0.006
> 65	390 (51.2%)	438 (58.3%)	
Sex			0.533
Male	482 (63.3%)	464 (61.8%)	
Female	279 (36.7%)	287 (38.2%)	
ECOG-PS			< 0.001
0/1	657 (86.3%)	580 (77.2%)	
2/3/4	104 (13.7%)	171 (22.8%)	
Smoking			< 0.001
No	157 (20.6%)	573 (76.3%)	
Yes	42 (5.5%)	158 (21.0%)	
Unknown	562 (73.9%)	20 (2.7%)	
Histology subtypes			0.009
Adenocarcinoma	728 (95.7%)	736 (98.0%)	
Non-adenocarcinoma	33 (4.3%)	15 (2.0%)	
EGFR mutation subtypes			0.956
Exon 19 deletion	342 (44.9%)	340 (45.3%)	
L858R	379 (49.8%)	372 (49.5%)	
Compound	16 (2.1%)	18 (2.4%)	
Major uncommon	24 (3.2%)	21 (2.8%)	
Stage			0.863
III	49 (6.4%)	50 (6.7%)	
IV	712 (93.6%)	701 (93.3%)	
Liver metastasis			0.467
Yes	87 (11.4%)	95 (12.6%)	
No	674 (88.6%)	656 (87.4%)	
Brain metastasis			0.613
Yes	233 (30.6%)	239 (31.8%)	
No	528 (9.4%)	512 (68.2%)	

Abbreviations: ECOG, Eastern Cooperative Oncology Group; EGFR, epidermal growth factor receptor; PS, performance score; SD, standard deviation.

mance was based on two main aspects: calibration and discrimination. First, the calibration curve was plotted to compare the nomogram-predicted versus observed probability of survival, which was conducted by bootstrapping through 1,000 bootstrap resamples internally with the NCKUH derivation cohort and externally with the CGMH validation cohort [14]. Second, the discrimination measures of the prediction model were evaluated using Harrell's C-index [15], Gonen and Heller's K statistic [16], and Royston and Sauerbrei's D [17]. Furthermore, we investigated the discriminative ability for risk stratification according to the linear predictor derived from the nomogram using recursive partitioning analysis (RPA), a statistical methodology that creates a survival analysis tree and establishes an optimal cut-off point that better predicts disease progression [18]. These discrimination factors refer to the nomogram model's ability to correctly distinguish the outcomes among different patient subgroups by Kaplan-Meier analysis and compared with the log-rank test. Statistical analyses were conducted using the SPSS software (IBM Corp. Released 2011. IBM SPSS Statistics

Table 2. Univariate analysis of prognostic factors for progression-free survival

Parameter	Total N	N of Events (%)	Median (Months)	95% CI	*p*-Value	Hazard Ratio	95% CI	*p*-Value
Age (years)					0.357			
≤ 65	371	342 (92.2)	11.7	10.9-12.5		1		
> 65	390	353 (90.5)	12.4	10.9-13.9		0.932	0.804-1.082	0.357
Sex					0.015			
Male	279	258 (92.5)	10.8	9.5-12.2		1.210	1.037-1.412	0.015
Female	482	437 (90.7)	12.4	11.2-13.7		1		
ECOG-PS					< 0.001			
0/1	657	596 (90.7)	12.3	11.2-13.3		1		
2/3/4	104	99 (95.2)	7.8	6.0-9.5		1.692	1.366-2.096	< 0.001
Smoking					0.174			
No	157	144 (91.7%)	13.3	11.2-15.5		1		
Yes	42	40 (95.2%)	10.8	7.9-13.8		1.228	0.865-1.745	0.250
Unknown	562	511 (90.9%)	11.8	10.9-12.7		1.188	0.986-1.432	0.070
Histology subtypes					0.002			
Adenocarcinoma	728	663 (91.1)	12.2	11.3-13.0		1		
Non-adenocarcinoma	33	32 (97.0)	6.8	4.1-9.6		1.739	1.219-2.482	0.002
EGFR mutation subtypes					0.042			
Exon 19 deletion	342	311 (90.9)	12.0	10.9-13.1		0.552	0.361-0.843	0.006
L858R	379	348 (91.8)	12.2	10.8-13.6		0.549	0.360-0.838	0.005
Compound	16	13 (81.2)	13.3	9.5-17.0		0.523	0.265-1.035	0.063
Major uncommon	24	23 (95.8)	5.9	4.7-7.1		1		
Stage					< 0.001			
III	49	36 (73.5)	29.6	16.7-42.5		1		
IV	712	659 (92.6)	11.8	11.1-12.5		2.154	1.536-3.022	< 0.001
Liver metastasis					< 0.001			
Yes	87	85 (97.7)	8.4	7.8-9.1		2.179	1.726-2.751	< 0.001
No	674	610 (90.5)	12.5	11.4-13.6		1		
Brain metastasis					< 0.001			
Yes	233	219 (94.0)	10.7	9.5-12.0		1.369	1.165-1.609	< 0.001
No	528	476 (90.2)	12.6	11.3-13.8		1		

Abbreviations: CI, confidence interval; ECOG, Eastern Cooperative Oncology Group; EGFR, epidermal growth factor receptor; PS, performance score.

for Windows, version 20.0. Armonk, NY, USA: IBM Corp.). All *p* values were two-sided, and statistical significance was set at P < 0.05.

Results

Patient characteristics

After enrollment, 761 patients with treatment-naïve EGFR-mutant advanced-stage NSCLC were enrolled in the derivation cohort, and 751 patients with treatment-naïve EGFR-mutant advanced-stage NSCLC from CGMH were enrolled in the validation cohort. As summarized in **Table 1**, the baseline clinical characteristics of the two groups were comparable. The patients in the derivation cohort were significantly younger (65.8 ± 11.6 versus 68.1 ± 11.8, P < 0.001) and had better performance status than those in the validation cohort. Most patients in the present study had histologically confirmed lung adenocarcinoma. Other characteristics, including sex, EGFR mutation subtypes, stage, presence of liver metastasis, and presence of brain metastasis, were comparable between the two cohorts.

Influence of clinical variables on PFS

Univariate analysis of the possible prognostic factors revealed that PFS was statistically different among patients of different sexes (male

Table 3. Multivariate analysis of prognostic variables associated with progression-free survival

Prognostic Variable	Adjusted Hazard Ratio	95% CI		*p*-Value	Points Assigned in Nomogram
		Lower	Upper		
Sex					
Male	1.228	1.051	1.435	0.010	32
Female	1				0
ECOG-PS					
0/1	1				0
2/3/4	1.463	1.174	1.825	< 0.001	59
Histology subtypes					
Adenocarcinoma	1				0
Non-adenocarcinoma	1.741	1.215	2.493	0.002	85
EGFR mutation subtypes					
Exon 19 deletion	0.591	0.384	0.908	0.016	18
L858R	0.594	0.387	0.911	0.017	18
Compound	0.527	0.266	1.045	0.066	0
Major uncommon	1				99
Stage					
IIIB	1				0
IV	1.906	1.351	2.688	< 0.001	99
Liver metastasis					
Yes	1.915	1.509	2.430	< 0.001	100
No	1				0
Brain metastasis					
Yes	1.201	1.015	1.420	0.033	28
No	1				0

Abbreviations: CI, confidence interval; ECOG, Eastern Cooperative Oncology Group; EGFR, epidermal growth factor receptor; PS, performance score.

vs. female, P = 0.015), PS (0/1 vs. 2/3/4, P < 0.001), histology subtypes (adenocarcinoma vs. non-adenocarcinoma, P = 0.002), EGFR mutation subtypes (L858R vs. exon 19 deletion vs. compound vs. major uncommon, P = 0.042), stage (IIIB vs. IV, P < 0.001), presence of liver metastasis (P < 0.001), and presence of brain metastasis (P < 0.001) (**Table 2**).

After incorporation of all prognostic factors into the univariate Cox proportional hazard regression models, seven variables were selected as prognostic factors, including sex (male vs. female, adjusted hazard ratio [AHR]: 1.210, 95% confidence interval [CI]: 1.037-1.412, P = 0.015), PS (2/3/4 vs. 0/1, AHR: 1.692, 95% CI: 1.366-2.096, P < 0.001), histology subtype (non-adenocarcinoma vs. adenocarcinoma, AHR: 1.739, 95% CI: 1.219-2.482, P = 0.002), EGFR mutation subtype (exon 19 deletion vs. major uncommon, AHR: 0.552, 95% CI: 0.361-0.843, P = 0.006; L858R vs. major uncommon, AHR: 0.549, 95% CI: 0.360-0.838, P = 0.005; compound vs. major uncommon, AHR: 0.523, 95% CI: 0.265-1.035, P = 0.063), stage (IV vs. IIIB, AHR: 2.154, 95% CI: 1.536-3.022, P < 0.001), presence of liver metastasis (AHR: 2.179, 95% CI: 1.726-2.751, P < 0.001), and presence of brain metastasis (AHR: 1.369, 95% CI: 1.165-1.609, P < 0.001) (**Table 2**).

Establishment of a prognostic nomogram based on pretreatment variables

After incorporating all the seven significant prognostic factors into the multivariate analysis, seven variables were found to be independent prognostic factors for PFS (**Table 3**). To construct the prognostic nomogram, seven predictors (P < 0.05) derived from the multivariate analysis were incorporated to establish a nomogram (**Figure 1**). The risk score assigned for each predictor is shown in **Table 3**, including sex (male: 32 points; female: 0 point), ECOG-PS

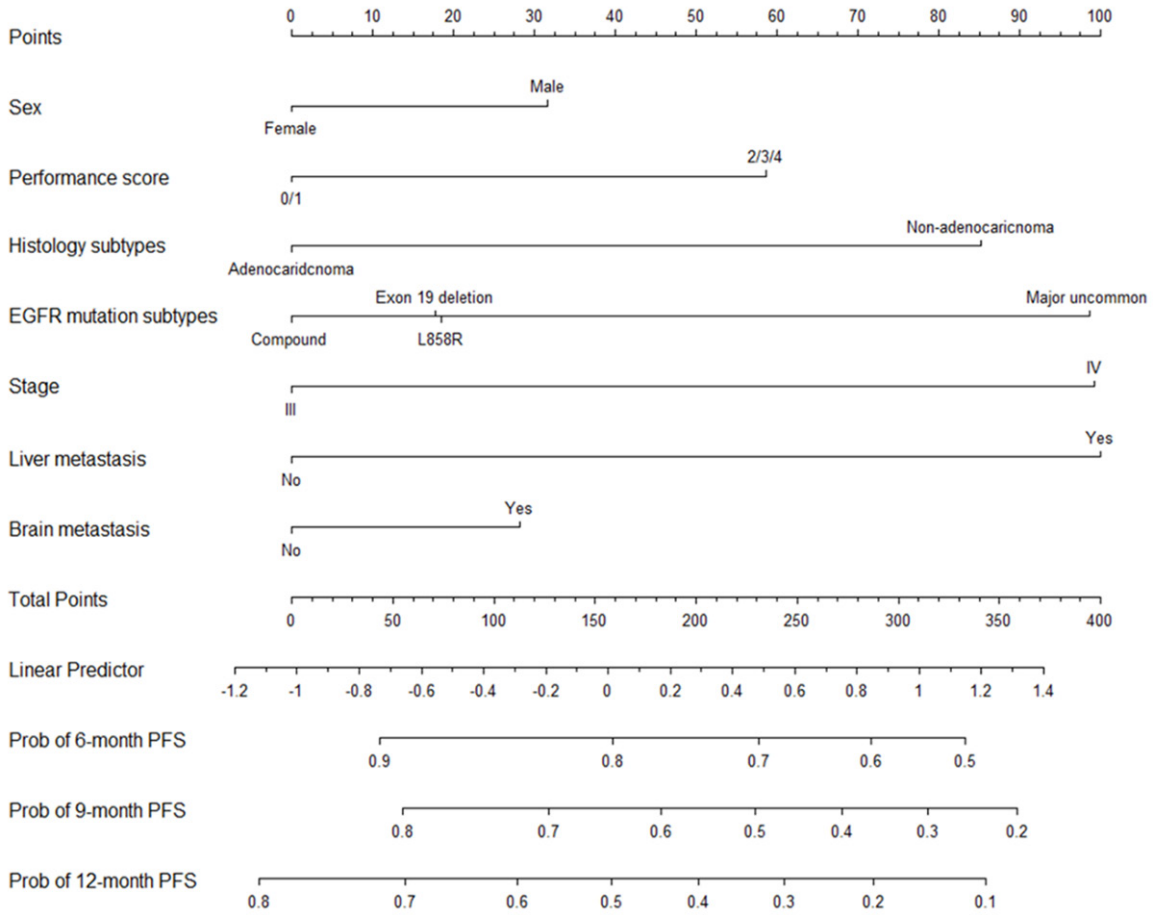


Figure 1. Predictive Nomogram for PFS. Nomogram based on the probability of progression-free survival using the Cox proportional hazard regression model from 761 patients with EGFR mutation-positive non-small cell lung cancer. EGFR, epidermal growth factor receptor; PFS, progression-free survival.

(0-1: 0 point; 2-4: 59 points), histologic subtype (adenocarcinoma: 0 point; non-adenocarcinoma: 85 points), EGFR mutation subtype (exon 19 deletion: 18 points; L858R: 18 points; compound: 0 point; major uncommon: 99 points), stage (IIIB: 0 point; IV: 99 points), presence of liver metastasis (100 points), and presence of brain metastasis (28 points). The total risk scores of the individuals ranged from 0 to 538.

The 6-, 9-, and 12-month PFS were estimated according to the linear predictor based on nomogram-derived total points (**Table 4** and **Figure 1**). Higher nomogram scores were associated with lower PFS rates. The calibration plots presented an excellent agreement for the 6-, 9-, and 12-year PFS in both the derivation (**Figure 2A-C**) and validation cohorts (**Figure 2D-F**). The C-index was 0.611 (95% CI: 0.600-0.622) and 0.606 (95% CI: 0.593-0.619) when comparing nomogram-predicted outcomes with the actual observed outcomes in the derivation and validation cohorts, respectively (**Table 5**). Other discrimination measures, including Gonen and Heller's K (0.597 vs. 0.590) and Royston and Sauerbrei's D (0.608 vs. 0.548), were also similar between the derivation and validation cohorts (**Table 5**).

Risk stratification by the nomogram for PFS

Based on the identified multiple variables, we used RPA to classify patients into five homogenous prognostic groups through a linear predictor for the best prediction of the risk of tumor progression (**Figure 3**). However, upon further analysis, we found that three of these groups had similar survival outcomes. As a result, we merged the three groups into one, resulting in a final classification of three homogenous prognostic groups: low-risk, intermediate-risk, and high-risk (**Figure 3**). The Kaplan-Meier curve for

Table 4. The prognostic scoring system

Nomogram Points	Probability of 6-Month PFS
333	0.50
286	0.60
231	0.70
159	0.80
43	0.90
Nomogram Points	Probability of 9-Month PFS
359	0.20
314	0.30
272	0.40
229	0.50
182	0.60
127	0.70
55	0.80
Nomogram Points	Probability of 12-Month PFS
343	0.10
288	0.20
243	0.30
201	0.40
158	0.50
111	0.60
56	0.70

Abbreviations: PFS, progression-free survival.

PFS according to risk subgroups in the NCKUH derivation cohort is shown in **Figure 4**, demonstrating that patients in different risk subgroups had significant differences in the risk of disease progression (**Figure 4** and **Table 5**). Finally, the cut-off point of the linear predictor was transformed into the relevant nomogram scores. A total of 167 patients were categorized into the high-risk group (total points: > 202, median PFS: 8.0 months), 550 into the intermediate-risk group (total points: 79-202, median PFS: 13.1 months), and 44 into the low-risk group (total points: 0-78, median PFS: 30.2 months) (**Figure 4**). This result was further validated using the CGMH cohort. The PFS analyzed using the Kaplan-Meier method also revealed significantly different prognoses among patients who received first-line EGFR-TKIs (**Figure 5** and **Table 5**).

Risk stratification by the nomogram for OS

Adopting a similar approach to that used to establish the nomogram for PFS, we identified seven prognostic factors for OS through univariate and multivariate analyses (Supplementary Tables 1, 2). Subsequently, we built a prognostic nomogram for OS (Supplementary Table 3 and Supplementary Figure 1) that exhibited strong model performance, as evidenced by discrimination measures, such as Harrell's C-index, Gonen and Heller's K, and Royston & Sauerbrei's D (Supplementary Table 4). Additionally, the calibration plots showed accurate calibration of the nomogram (Supplementary Figure 2). Using RPA, we categorized patients into four distinct prognostic groups based on a linear predictor for the most accurate estimation of death risk (Supplementary Figure 3).

Supplementary Figure 4 illustrates the Kaplan-Meier curve for OS, stratified by risk subgroups within the NCKUH derivation cohort. The curve shows notable disparities in death risk among patients from various risk subgroups (Supplementary Figure 4 and Supplementary Table 4). The cut-off point for the linear predictor was transformed into the relevant nomogram scores. A total of 197 patients were categorized into the very high-risk group (total points: > 136, median OS: 14.9 months), 47 into the high-risk group (total points: 112-136, median OS: 19.7 months), 502 into the intermediate-risk group (total points: 31-112, median OS: 33.4 months), and 15 into the low-risk group (total points: 0-31, median OS: 85.2 months) (Supplementary Figure 4). The robustness of these findings was validated in the CGMH cohort (Supplementary Figure 5 and Supplementary Table 4).

Discussion

In the present study, a nomogram based on pretreatment characteristics was established to predict PFS among patients with EGFR-mutant advanced-stage NSCLC who received first-line EGFR-TKIs. Eight parameters were used: sex, ECOG-PS, histology subtypes, EGFR mutation subtypes, stage, choice of first-line EGFR-TKIs, presence of liver metastasis, and presence of brain metastasis. This nomogram can be easily used by clinicians to predict the 6-, 9-, and 12-month PFS rates and stratify patients into different risk groups. In addition to internal bootstrapping validation, this nomogram performed well in external validation, indicating that this stratification could help clinicians better evaluate the optimal therapeutic options before the initiation of EGFR-TKIs.

Different metastatic sites may confer distinct prognoses and influence the response to EGFR

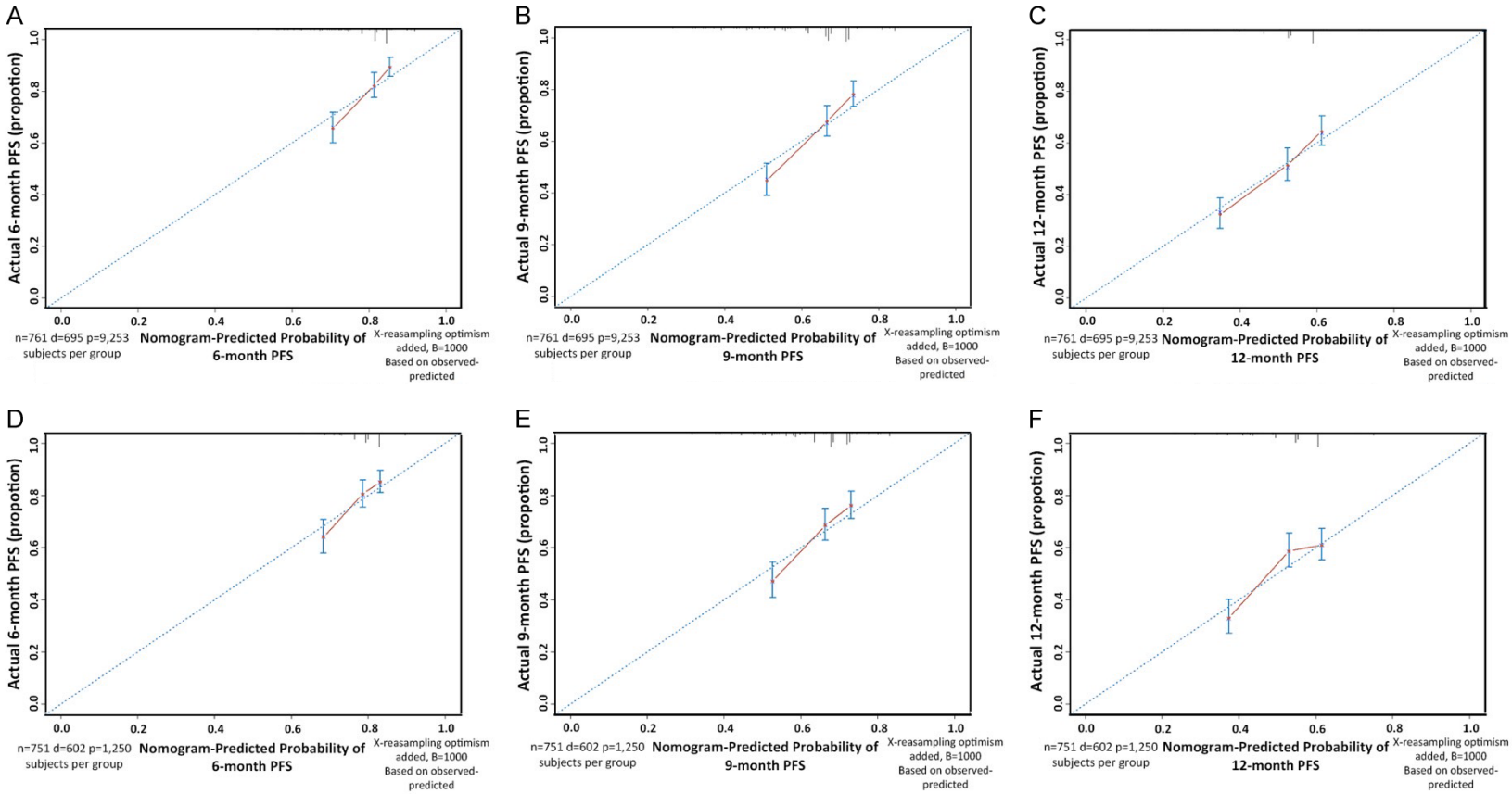


Figure 2. Internal and external calibration curves. These show nomogram-predicted and actually observed PFS. A. Internal calibration for 6-month PFS. B. Internal calibration for 9-month PFS. C. Internal calibration for 12-month PFS. D. External calibration for 6-month PFS. E. External calibration for 9-month PFS. F. External calibration for 12-month PFS. PFS, progression-free survival.

Table 5. Discrimination measures and hazard ratios evaluated between two datasets

Measures of discrimination	NCKUH derivation cohort			CGMH validation cohort		
	Estimate	SE		Estimate	SE	
Harrell's C-index	0.611	0.011		0.606	0.013	
Gonen and Heller's K	0.597	0.010		0.590	0.010	
Royston & Sauerbrei's D	0.608	0.065		0.548	0.069	
Prognostic groups[a]	Hazard ratio	95% CI of HR	*p*-Value	Hazard ratio	95% CI of HR	*p*-Value
Low risk[b]	1			1		
Intermediate risk	2.099	1.456-3.026	< 0.001	1.815	1.235-2.666	0.002
High risk	4.412	2.978-6.536	< 0.001	3.424	2.285-5.132	< 0.001

Abbreviations: CI, confidence interval; HR, hazard ratio; SE, standard error. [a]NCKUH Prognostic group was categorized using recursive partitioning analysis (RPA) to establish optimal cutoff points by Cox's model. [b]reference category.

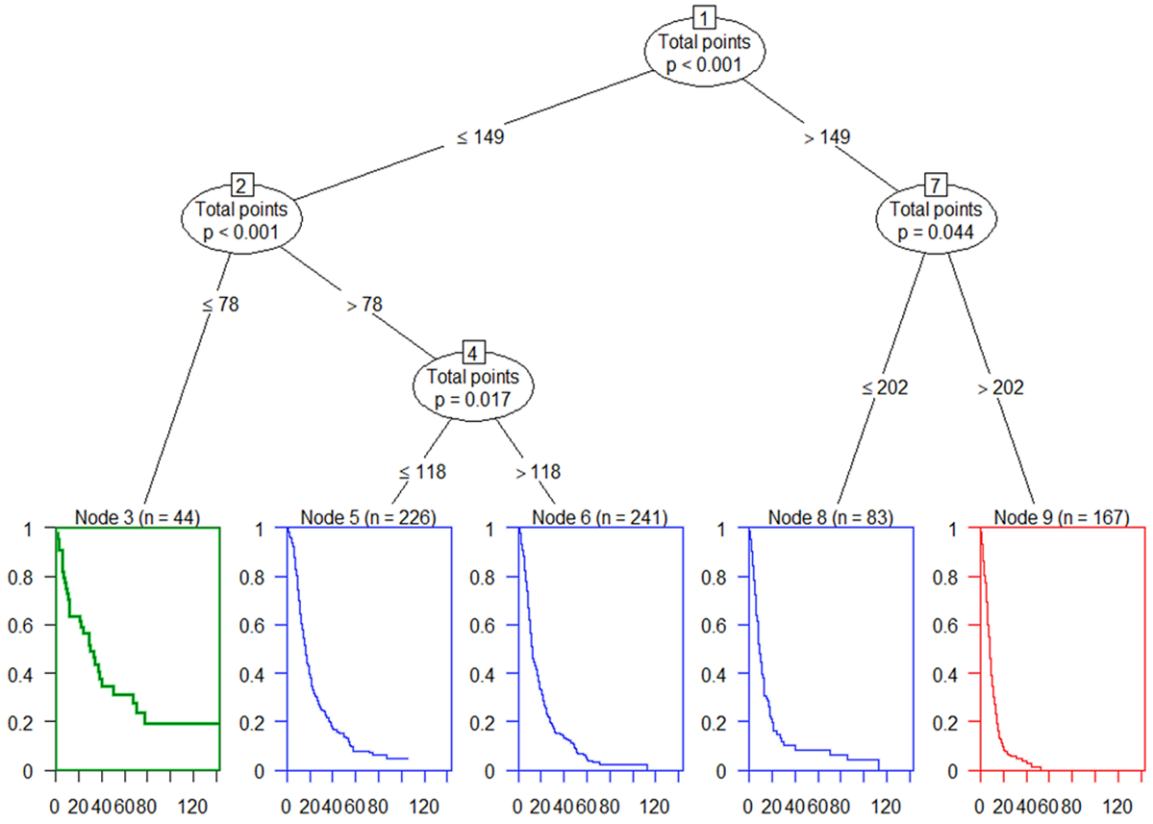


Figure 3. Survival analysis tree. A survival analysis tree was used to establish an optimal cutoff point to better predict disease progression.

TKI. Brain metastasis is a well-known poor prognostic factor in patients with NSCLC [19]. It was also a predictor of a higher risk in our prediction model. Liver involvement appears in 10-20% of the patients and is a significant poor prognostic factor for patients with EGFR-mutant NSCLC [20, 21]. The hepatic immune microenvironment could promote the survival of lung adenocarcinoma cells via the METTL3-activated YAP1/TEAD signaling pathway [22], which might be associated with drug resistance to EGFR-TKIs. Consistent with these studies, the presence of liver metastasis was the strongest variable with the highest scores (100 points) in our prediction model and was associated with shorter PFS when receiving first-line EGFR-TKIs [23].

Unlike common EGFR mutations, uncommon EGFR mutations, such as G719X, S768I, or

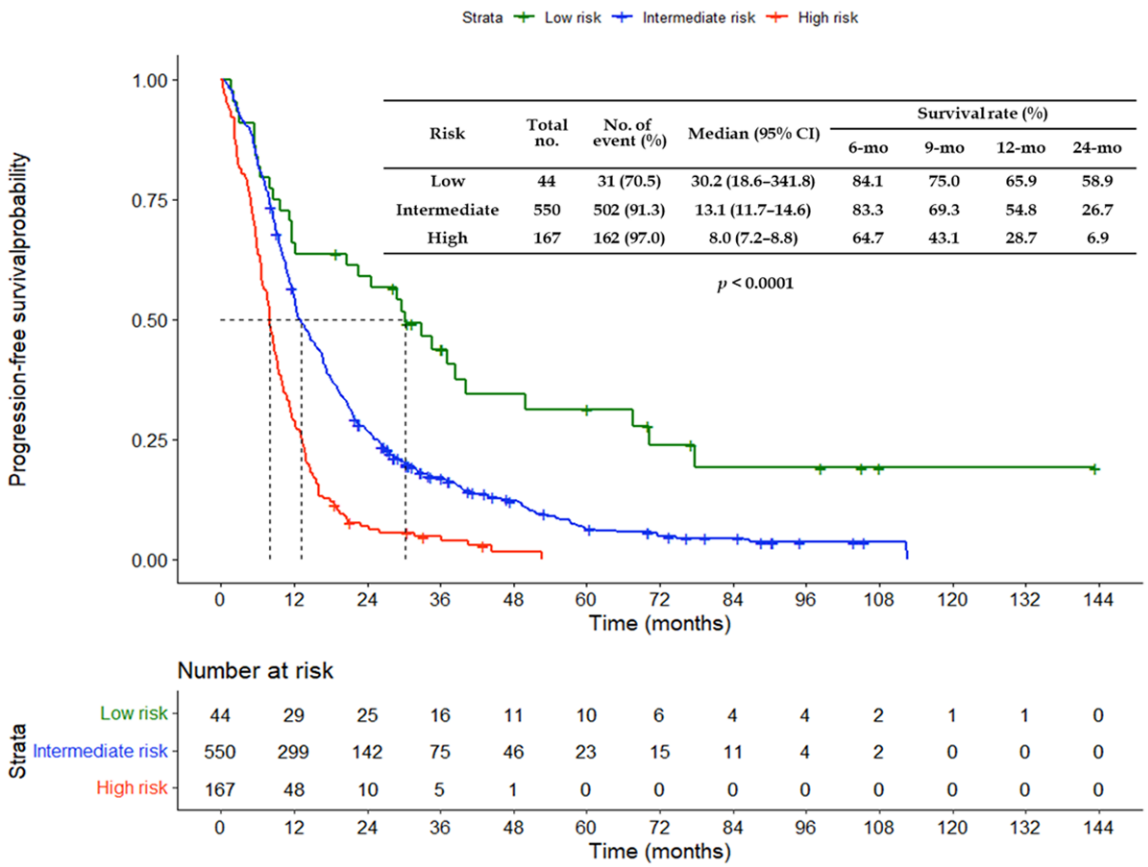


Figure 4. Kaplan-Meier curve of progression-free survival (PFS) in NCKUH cohort. Kaplan-Meier curve for PFS in the NCKUH derivation cohort according to risk groups. CI, confidence interval.

L861Q, would cause P-loop and αC-helix compression of the ATP-binding pocket in the kinase domain of EGFR, which affects the binding affinity of first-generation EGFR-TKIs, deteriorating its efficacy [24]. In our study, the presence of a major uncommon mutation was an important poor prognostic factor (94 points, third highest), which is consistent with the results of previous studies. In a subgroup of NSCLC patients with uncommon EGFR mutations, afatinib demonstrated promising PFS in a combined post-hoc analysis of LUX-Lung 2, LUX-Lung 3, and LUX-Lung 6 [25]. A subsequent real-world study further revealed that afatinib could achieve a superior PFS compared with gefitinib or erlotinib [26-28]. However, a previous study that used a nomogram to predict PFS did not include uncommon EGFR mutations [13, 29]. To the best of our knowledge, our nomogram is the first prediction model to include patients with uncommon EGFR mutations.

In recent decades, a growing number of studies have established nomograms for predicting survival outcomes in patients with EGFR-mutant NSCLC. Keam et al. developed a nomogram to predict survival outcomes in patients with EGFR-mutant NSCLC [30]. However, not all enrolled patients received first-line EGFR-TKIs. In addition, the prediction model included tumor response as a prognostic factor, which is difficult to detect before the initiation of EGFR-TKIs [30]. Chen et al. collected data from 13,043 patients with advanced adenocarcinoma to develop a nomogram for predicting 1-year and 2-year survival probabilities [31]. Although the sample size was large, the study included patients with and without EGFR mutations. Furthermore, the prognostic factors selected in patients with EGFR-mutant NSCLC did not specify the subtype of EGFR mutation or the choice of first-line EGFR-TKIs [31]. Another study by Chang et al. created a nomogram that included 2190 EGFR-mutant NSCLC

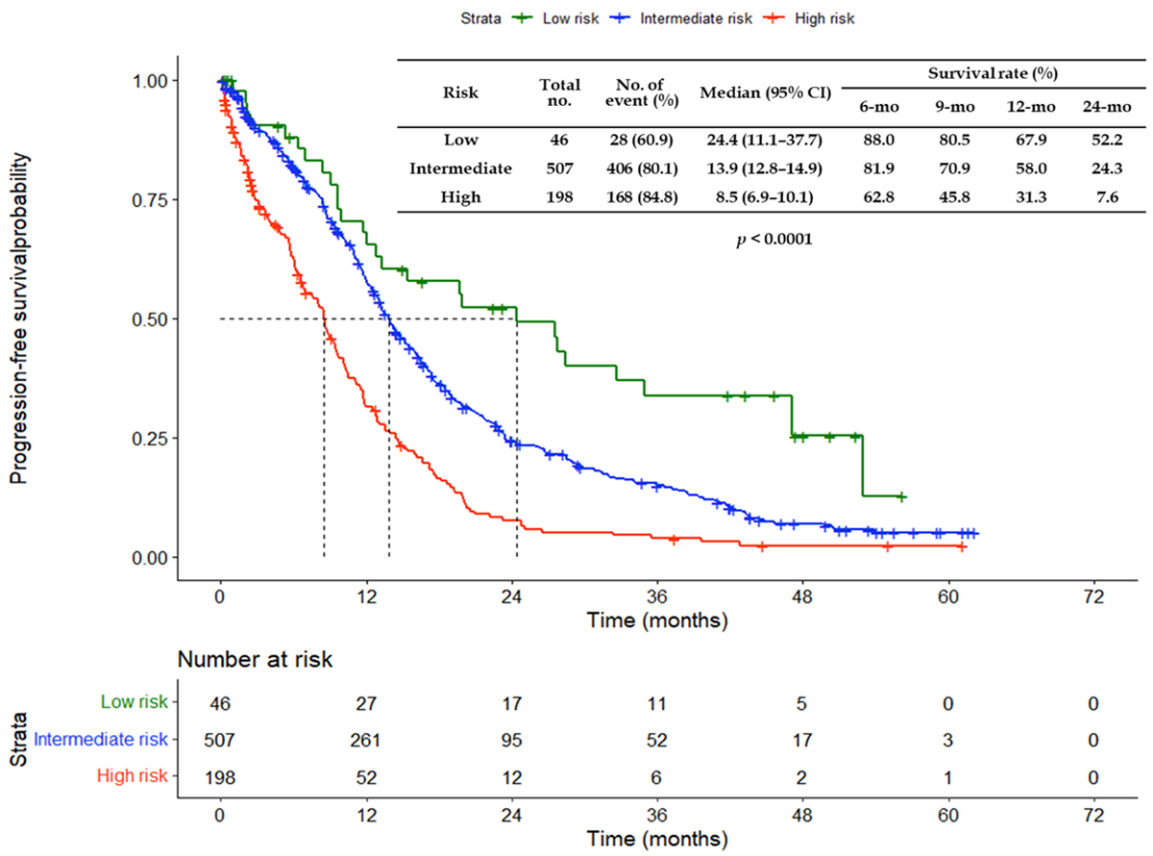


Figure 5. Kaplan-Meier curve of progression-free survival (PFS) in CGMH cohort. Kaplan-Meier curve for PFS in the CGMH validation cohort according to risk groups. CI, confidence interval.

patients and had good performance in risk stratification [13], but rare EGFR mutations were not included in the prediction model, and the model did not undergo external validation. In the present study, we established a nomogram that comprised both common and uncommon EGFR mutations and showed good performance after internal and external validation. Moreover, this study represents the first attempt to establish a nomogram for predicting OS in patients with metastatic NSCLC harboring EGFR mutations.

Our study has several limitations. First, a nomogram was built based on retrospective analysis. However, all prognostic variables included in the present study were objective outcomes. Regarding the survival outcome, all patients received regular imaging examinations based on the reimbursement criteria in Taiwan, which could detect disease progression in a time-dependent manner. Second, the third-generation EGFR-TKI osimertinib was not included in this study. Although the FLAURA trial demonstrated better PFS and OS for osimertinib compared with first-generation EGFR-TKIs [8, 9], the OS benefit was insignificant in the Asian population. In addition, no prospective trials have compared the treatment efficacy of second- and third-generation EGFR-TKIs. Furthermore, the real-world study Giotag [10] demonstrated excellent outcomes of sequential EGFR-TKI use, and another CJLSG1903 study [32] revealed that afatinib performed better in certain subgroups. First-line therapy with first- or second-generation EGFR-TKIs still plays an important role in the Asian population.

In conclusion, we established a nomogram with external validation for predicting PFS among patients with EGFR-mutant NSCLC based on pretreatment clinical characteristics. This nomogram could help clinicians predict the risk of progression and administer individualized treatment regimens to improve survival outcomes.

Acknowledgements

This study was based in part on data sourced from the Cancer Data Bank of the National Cheng Kung University Hospital and the Chang Gung Memorial Hospital. This study was funded by the National Science and Technology Council of Taiwan (Grant No. NSTC 109-2314-B-182-080-MY3, NSTC 110-2314-B-006-098-MY3, NSTC 111-2811-B-182-017, and NSTC 111-2314-B-006-092-MY3), Linkou Chang Gung Memorial Hospital (Grant No. CMRPG3M0971), and the National Cheng Kung University Hospital (Grant No. NCKUH-1120-2051 and NCKUH-11203021).

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Cheng-Ta Yang, Division of Thoracic Oncology, Department of Thoracic Medicine, Linkou Chang Gung Memorial Hospital, College of Medicine, Chang Gung University, No. 5, Fuxing St., Guishan Dist., Taoyuan 333, Taiwan. E-mail: yang1946@cgmh.org.tw; Dr. Chiao-En Wu, Division of Hematology-Oncology, Department of Internal Medicine, Linkou Chang Gung Memorial Hospital, College of Medicine, Chang Gung University, No. 5, Fuxing St., Guishan Dist., Taoyuan 333, Taiwan. E-mail: 8805017@cgmh.org.tw

References

[1] Tan AC and Tan DSW. Targeted therapies for lung cancer patients with oncogenic driver molecular alterations. J Clin Oncol 2022; 40: 611-625.

[2] Mok TS, Wu YL, Thongprasert S, Yang CH, Chu DT, Saijo N, Sunpaweravong P, Han B, Margono B, Ichinose Y, Nishiwaki Y, Ohe Y, Yang JJ, Chewaskulyong B, Jiang H, Duffield EL, Watkins CL, Armour AA and Fukuoka M. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. N Engl J Med 2009; 361: 947-957.

[3] Inoue A, Kobayashi K, Maemondo M, Sugawara S, Oizumi S, Isobe H, Gemma A, Harada M, Yoshizawa H, Kinoshita I, Fujita Y, Okinaga S, Hirano H, Yoshimori K, Harada T, Saijo Y, Hagiwara K, Morita S and Nukiwa T; North-East Japan Study Group. Updated overall survival results from a randomized phase III trial comparing gefitinib with carboplatin-paclitaxel for chemo-naïve non-small cell lung cancer with sensitive EGFR gene mutations (NEJ002). Ann Oncol 2013; 24: 54-59.

[4] Zhou C, Wu YL, Chen G, Feng J, Liu XQ, Wang C, Zhang S, Wang J, Zhou S, Ren S, Lu S, Zhang L, Hu C, Hu C, Luo Y, Chen L, Ye M, Huang J, Zhi X, Zhang Y, Xiu Q, Ma J, Zhang L and You C. Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. Lancet Oncol 2011; 12: 735-742.

[5] Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E, Palmero R, Garcia-Gomez R, Pallares C, Sanchez JM, Porta R, Cobo M, Garrido P, Longo F, Moran T, Insa A, De Marinis F, Corre R, Bover I, Illiano A, Dansin E, de Castro J, Milella M, Reguart N, Altavilla G, Jimenez U, Provencio M, Moreno MA, Terrasa J, Muñoz-Langa J, Valdivia J, Isla D, Domine M, Molinier O, Mazieres J, Baize N, Garcia-Campelo R, Robinet G, Rodriguez-Abreu D, Lopez-Vivanco G, Gebbia V, Ferrera-Delgado L, Bombaron P, Bernabe R, Bearz A, Artal A, Cortesi E, Rolfo C, Sanchez-Ronco M, Drozdowskyj A, Queralt C, de Aguirre I, Ramirez JL, Sanchez JJ, Molina MA, Taron M and Paz-Ares L; Spanish Lung Cancer Group in collaboration with Groupe Français de Pneumo-Cancérologie and Associazione Italiana Oncologia Toracica. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. Lancet Oncol 2012; 13: 239-246.

[6] Wu YL, Zhou C, Liam CK, Wu G, Liu X, Zhong Z, Lu S, Cheng Y, Han B, Chen L, Huang C, Qin S, Zhu Y, Pan H, Liang H, Li E, Jiang G, How SH, Fernando MCL, Zhang Y, Xia F and Zuo Y. First-line erlotinib versus gemcitabine/cisplatin in patients with advanced EGFR mutation-positive non-small-cell lung cancer: analyses from the phase III, randomized, open-label, ENSURE study. Ann Oncol 2015; 26: 1883-1889.

[7] Mitsudomi T, Morita S, Yatabe Y, Negoro S, Okamoto I, Tsurutani J, Seto T, Satouchi M, Tada H, Hirashima T, Asami K, Katakami N, Takada M, Yoshioka H, Shibata K, Kudoh S, Shimizu E, Saito H, Toyooka S, Nakagawa K and Fukuoka M; West Japan Oncology Group. Gefitinib versus cisplatin plus docetaxel in patients with non-small-cell lung cancer harbouring mutations of the epidermal growth factor receptor (WJTOG3405): an open label, randomised phase 3 trial. Lancet Oncol 2010; 11: 121-128.

[8] Soria JC, Ohe Y, Vansteenkiste J, Reungwetwattana T, Chewaskulyong B, Lee KH, Dechaphunkul A, Imamura F, Nogami N, Kurata T, Okamoto I, Zhou C, Cho BC, Cheng Y, Cho EK, Voon

PJ, Planchard D, Su WC, Gray JE, Lee SM, Hodge R, Marotti M, Rukazenkov Y and Ramalingam SS; FLAURA Investigators. Osimertinib in untreated EGFR-mutated advanced non-small-cell lung cancer. N Engl J Med 2018; 378: 113-125.

[9] Ramalingam SS, Vansteenkiste J, Planchard D, Cho BC, Gray JE, Ohe Y, Zhou C, Reungwetwattana T, Cheng Y, Chewaskulyong B, Shah R, Cobo M, Lee KH, Cheema P, Tiseo M, John T, Lin MC, Imamura F, Kurata T, Todd A, Hodge R, Saggese M, Rukazenkov Y and Soria JC; FLAURA Investigators. Overall survival with osimertinib in untreated, EGFR-mutated advanced NSCLC. N Engl J Med 2020; 382: 41-50.

[10] Hochmair MJ, Morabito A, Hao D, Yang CT, Soo RA, Yang JC, Gucalp R, Halmos B, Märten A and Cufer T. Sequential afatinib and osimertinib in patients with EGFR mutation-positive non-small-cell lung cancer: final analysis of the GioTag study. Future Oncol 2020; 16: 2799-2808.

[11] Wu SG, Chiang CL, Liu CY, Wang CC, Su PL, Hsia TC, Shih JY and Chang GC. An observational study of acquired EGFR T790M-dependent resistance to EGFR-TKI treatment in lung adenocarcinoma patients in Taiwan. Front Oncol 2020; 10: 1481.

[12] Blakely CM, Watkins TBK, Wu W, Gini B, Chabon JJ, McCoach CE, McGranahan N, Wilson GA, Birkbak NJ, Olivas VR, Rotow J, Maynard A, Wang V, Gubens MA, Banks KC, Lanman RB, Caulin AF, St John J, Cordero AR, Giannikopoulos P, Simmons AD, Mack PC, Gandara DR, Husain H, Doebele RC, Riess JW, Diehn M, Swanton C and Bivona TG. Evolution and clinical impact of co-occurring genetic alterations in advanced-stage EGFR-mutant lung cancers. Nat Genet 2017; 49: 1693-1704.

[13] Chang JW, Huang CY, Fang YF, Chang CF, Yang CT, Kuo CS, Hsu PC and Wu CE. Risk stratification using a novel nomogram for 2190 EGFR-mutant NSCLC patients receiving the first or second generation EGFR-TKI. Cancers (Basel) 2022; 14: 977.

[14] Royston P and Altman DG. External validation of a Cox prognostic model: principles and methods. BMC Med Res Methodol 2013; 13: 33.

[15] Harrell FE Jr, Califf RM, Pryor DB, Lee KL and Rosati RA. Evaluating the yield of medical tests. JAMA 1982; 247: 2543-2546.

[16] Gönen M and Heller G. Concordance probability and discriminatory power in proportional hazards regression. Biometrika 2005; 92: 965-970.

[17] Royston P and Sauerbrei W. A new measure of prognostic separation in survival data. Stat Med 2004; 23: 723-748.

[18] Hothorn T, Hornik K and Zeileis A. Unbiased recursive partitioning: a conditional inference framework. J Comput Graph Stat 2006; 15: 651-674.

[19] Ernani V and Stinchcombe TE. Management of brain metastases in non-small-cell lung cancer. J Oncol Pract 2019; 15: 563-570.

[20] Chang YP, Chen YM, Lai CH, Lin CY, Fang WF, Huang CH, Li SH, Chen HC, Wang CC and Lin MC. The impact of de novo liver metastasis on clinical outcome in patients with advanced non-small-cell lung cancer. PLoS One 2017; 12: e0178676.

[21] Castañón E, Rolfo C, Viñal D, López I, Fusco JP, Santisteban M, Martin P, Zubiri L, Echeveste JI and Gil-Bazo I. Impact of epidermal growth factor receptor (EGFR) activating mutations and their targeted treatment in the prognosis of stage IV non-small cell lung cancer (NSCLC) patients harboring liver metastasis. J Transl Med 2015; 13: 257.

[22] Ni XF, Xie QQ, Zhao JM, Xu YJ, Ji M, Hu WW, Wu J and Wu CP. The hepatic microenvironment promotes lung adenocarcinoma cell proliferation, metastasis, and epithelial-mesenchymal transition via METTL3-mediated N6-methyladenosine modification of YAP1. Aging (Albany NY) 2021; 13: 4357-4369.

[23] Jiang T, Cheng R, Zhang G, Su C, Zhao C, Li X, Zhang J, Wu F, Chen X, Gao G, Li W, Cai W, Zhou F, Zhao J, Xiong A, Ren S, Zhang G, Zhou C and Zhang J. Characterization of liver metastasis and its effect on targeted therapy in EGFR-mutant NSCLC: a multicenter study. Clin Lung Cancer 2017; 18: 631-639, e632.

[24] Robichaux JP, Le X, Vijayan RSK, Hicks JK, Heeke S, Elamin YY, Lin HY, Udagawa H, Skoulidis F, Tran H, Varghese S, He J, Zhang F, Nilsson MB, Hu L, Poteete A, Rinsurongkawong W, Zhang X, Ren C, Liu X, Hong L, Zhang J, Diao L, Madison R, Schrock AB, Saam J, Raymond V, Fang B, Wang J, Ha MJ, Cross JB, Gray JE and Heymach JV. Structure-based classification predicts drug response in EGFR-mutant NSCLC. Nature 2021; 597: 732-737.

[25] Yang JC, Sequist LV, Geater SL, Tsai CM, Mok TS, Schuler M, Yamamoto N, Yu CJ, Ou SH, Zhou C, Massey D, Zazulina V and Wu YL. Clinical activity of afatinib in patients with advanced non-small-cell lung cancer harbouring uncommon EGFR mutations: a combined post-hoc analysis of LUX-Lung 2, LUX-Lung 3, and LUX-Lung 6. Lancet Oncol 2015; 16: 830-838.

[26] Kim Y, Lee SH, Ahn JS, Ahn MJ, Park K and Sun JM. Efficacy and safety of afatinib for EGFR-mutant non-small cell lung cancer, compared with gefitinib or erlotinib. Cancer Res Treat 2019; 51: 502-509.

[27] Shen YC, Tseng GC, Tu CY, Chen WC, Liao WC, Chen WC, Li CH, Chen HJ and Hsia TC. Comparing the effects of afatinib with gefitinib or Erlotinib in patients with advanced-stage lung adenocarcinoma harboring non-classical epidermal growth factor receptor mutations. Lung Cancer 2017; 110: 56-62.

[28] Chang JW, Huang CY, Fang YF, Chang CF, Yang CT, Kuo CS, Hsu PC and Wu CE. Epidermal growth factor receptor tyrosine kinase inhibitors for non-small cell lung cancer harboring uncommon EGFR mutations: real-world data from Taiwan. Thorac Cancer 2023; 14: 12-23.

[29] Du X, Bai H, Wang Z, Daun J, Liu Z, Xu J, Chang G, Zhu Y and Wang J. Establishment of prognostic nomograms for predicting the progression free survival of EGFR-sensitizing mutation, advanced lung cancer patients treated with EGFR-TKIs. Thorac Cancer 2022; 13: 1289-1298.

[30] Keam B, Kim DW, Park JH, Lee JO, Kim TM, Lee SH, Chung DH and Heo DS. Nomogram predicting clinical outcomes in non-small cell lung cancer patients treated with epidermal growth factor receptor tyrosine kinase inhibitors. Cancer Res Treat 2014; 46: 323-330.

[31] Chen HC, Tan EC, Liao CH, Lin ZZ and Yang MC. Development and validation of nomograms for predicting survival probability of patients with advanced adenocarcinoma in different EGFR mutation status. PLoS One 2019; 14: e0220730.

[32] Ito K, Morise M, Wakuda K, Hataji O, Shimokawaji T, Takahashi K, Furuya N, Takeyama Y, Goto Y, Abe T, Kato T, Ozone S, Ikeda S, Kogure Y, Yokoyama T, Kimura M, Yoshioka H, Murotani K, Kondo M and Saka H. A multicenter cohort study of osimertinib compared with afatinib as first-line treatment for EGFR-mutated non-small-cell lung cancer from practical dataset: CJLSG1903. ESMO Open 2021; 6: 100115.

Supplementary Table 1. Univariate analysis of prognostic factors for overall survival

Parameter	Total N	N of Events (%)	Median (Months)	95% CI	*p*-Value	Hazard Ratio	95% CI	*p*-Value
Age (years)					0.173			
≤ 65	371	288 (77.6)	28.4	25.1-31.6		1		
> 65	390	299 (76.7)	24.5	20.6-28.4		1.119	0.952-1.316	0.173
Sex					0.001			
Male	279	227 (81.4)	22.4	19.7-25.1		1.322	1.119-1.562	0.001
Female	482	360 (74.7)	30.0	26.8-33.2		1		
ECOG-PS					< 0.001			
0/1	657	491 (74.7)	30.0	27.3-32.7		1		
2/3/4	104	96 (92.3)	12.7	9.2-16.2		2.674	2.141-3.339	< 0.001
Smoking					0.060			
No	157	118 (75.2%)	31.2	22.7-39.8		1		
Yes	42	32 (76.2%)	24.5	16.0-33.0		1.606	0.717-1.567	0.771
Unknown	562	437 (77.8%)	26.1	23.3-28.9		1.267	1.032-1.555	0.024
Histology subtypes					< 0.001			
Adenocarcinoma	728	556 (76.4)	27.7	25.1-30.4		1		
Non-adenocarcinoma	33	31 (93.9)	12.8	6.3-19.4		2.114	1.471-3.038	< 0.001
EGFR mutation subtypes					0.180			
Exon 19 deletion	342	254 (74.3)	26.8	23.3-30.3		0.611	0.391-0.953	0.030
L858R	379	301 (79.4)	28.1	24.4-31.7		0.650	0.417-1.012	0.057
Compound	16	11 (68.7)	23.5	0.1-62.1		0.623	0.300-1.294	0.205
Major uncommon	24	21 (87.5)	15.2	7.4-23.0		1		
Stage					0.003			
III	49	30 (61.2)	45.7	22.8-68.7		1		
IV	712	557 (78.2)	26.1	23.6-28.6		1.749	1.211-2.528	0.003
Liver metastasis					< 0.001			
Yes	87	80 (92.0)	15.7	13.7-17.7		2.232	1.755-2.838	< 0.001
No	674	507 (75.2)	29.2	26.6-31.8		1		
Brain metastasis					0.001			
Yes	233	189 (81.1)	22.0	18.2-25.9		1.342	1.128-1.597	0.001
No	528	398 (75.4)	29.4	26.2-32.6		1		

Abbreviations: CI, confidence interval; ECOG, Eastern Cooperative Oncology Group; EGFR, epidermal growth factor receptor; PS, performance score.

Supplementary Table 2. Multivariate analysis of prognostic variables associated with overall survival

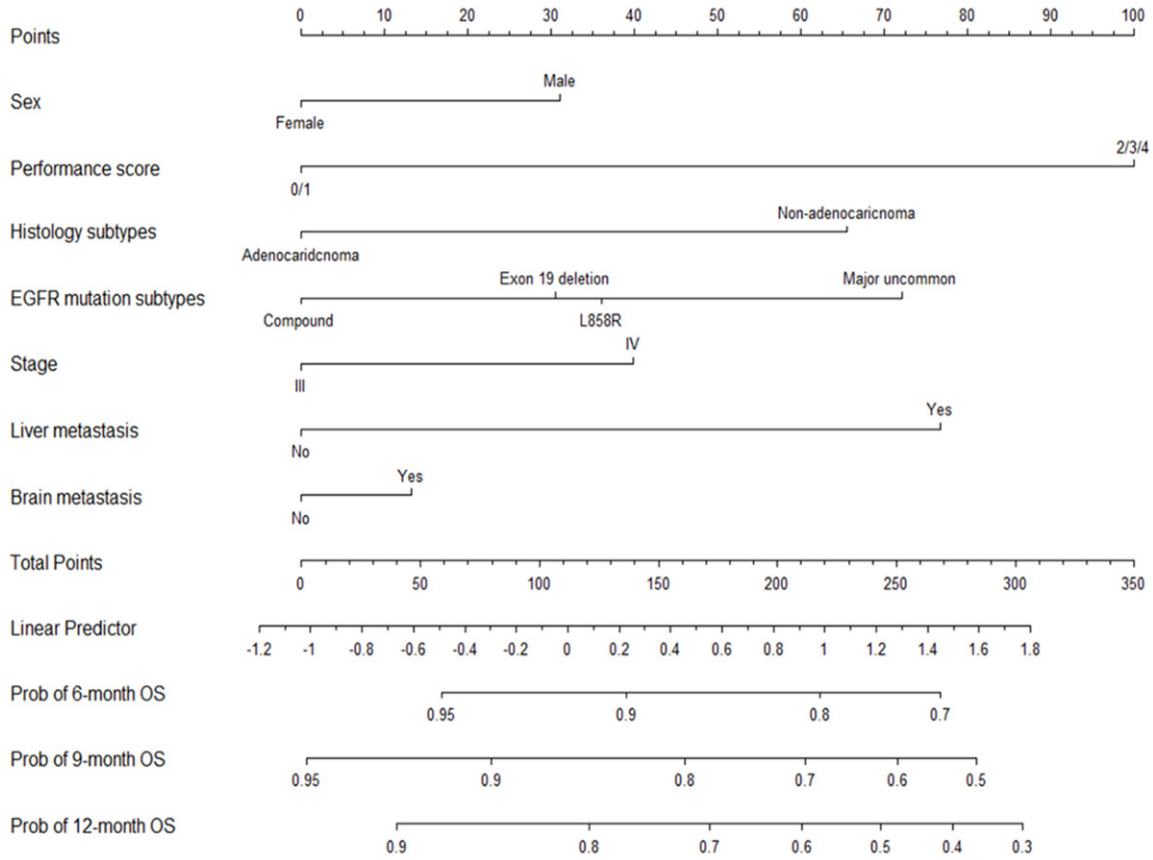
Prognostic Variable	Adjusted hazard ratio	95% CI		*p*-Value	Points Assigned in Nomogram
		Lower	Upper		
Sex					
Male	1.332	1.126	1.575	< 0.001	31
Female	1				0
ECOG-PS					
0/1	1				0
2/3/4	2.520	2.005	3.168	< 0.001	100
Histology subtypes					
Adenocarcinoma	1				0
Non-adenocarcinoma	1.833	1.271	2.642	0.001	66
EGFR mutation subtypes					
Exon 19 deletion	0.681	0.434	1.066	0.093	31
L858R	0.716	0.458	1.120	0.143	36
Compound	0.512	0.245	1.071	0.075	0
Major uncommon	1				72
Stage					
IIIB	1				0
IV	1.445	0.994	2.102	0.054	40
Liver metastasis					
Yes	2.034	1.588	2.603	< 0.001	77
No	1				0
Brain metastasis					
Yes	1.129	0.941	1.356	0.191	13
No	1				0

Abbreviations: CI, confidence interval; ECOG, Eastern Cooperative Oncology Group; EGFR, epidermal growth factor receptor; PS, performance score.

Supplementary Table 3. The prognostic scoring system

Nomogram Points	Probability of 6-Month OS
268	0.70
218	0.80
137	0.90
59	0.95
Nomogram Points	Probability of 9-Month OS
284	0.50
251	0.60
212	0.70
161	0.80
80	0.90
2	0.95
Nomogram Points	Probability of 12-Month OS
303	0.30
274	0.40
243	0.50
210	0.60
172	0.70
121	0.80
40	0.90

Abbreviations: OS, overall survival.

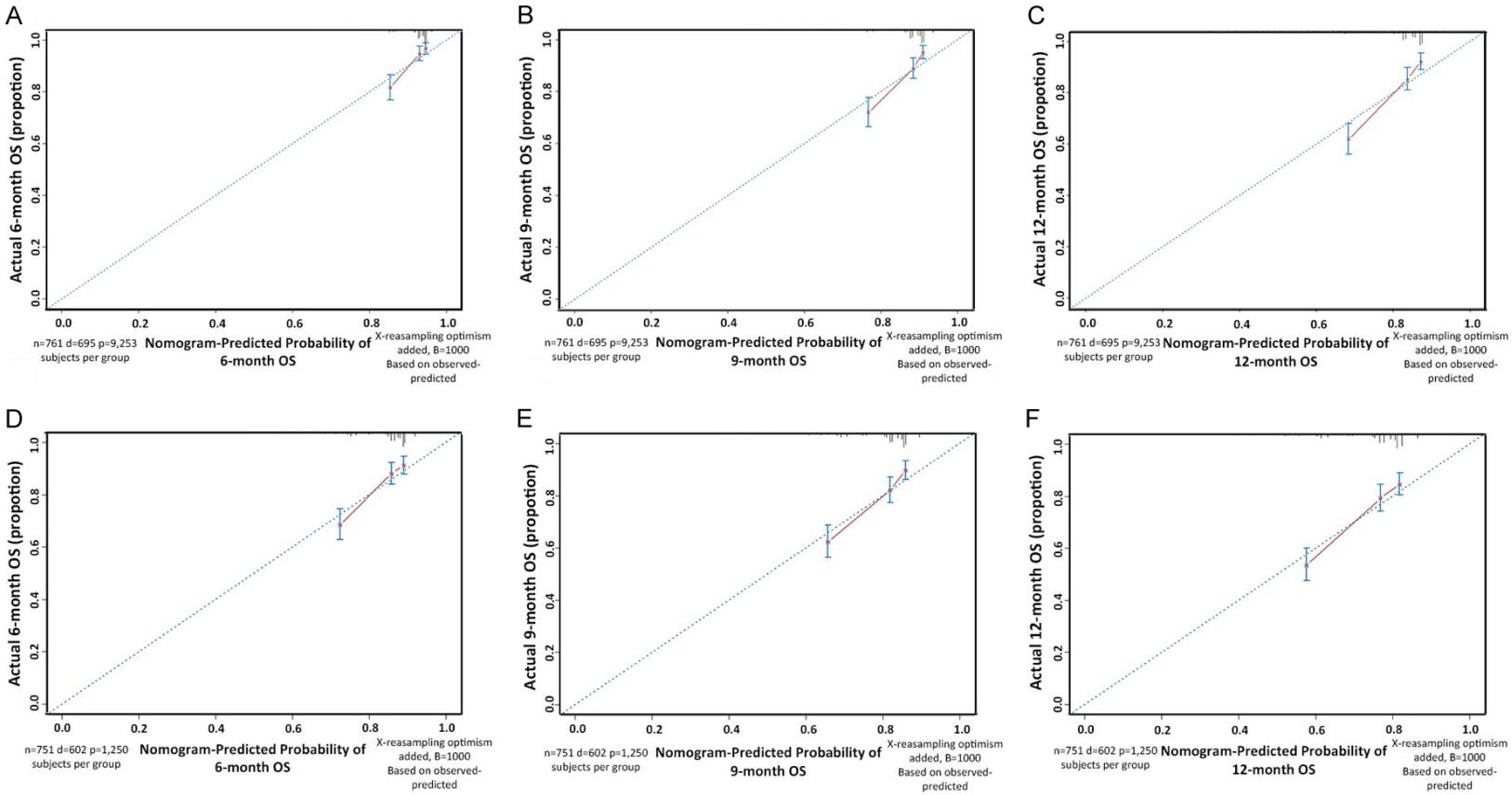


Supplementary Figure 1. Predictive Nomogram for overall survival (OS). Nomogram based on the probability of OS using the Cox proportional hazard regression model from 761 patients with EGFR mutation-positive non-small cell lung cancer. EGFR, epidermal growth factor receptor.

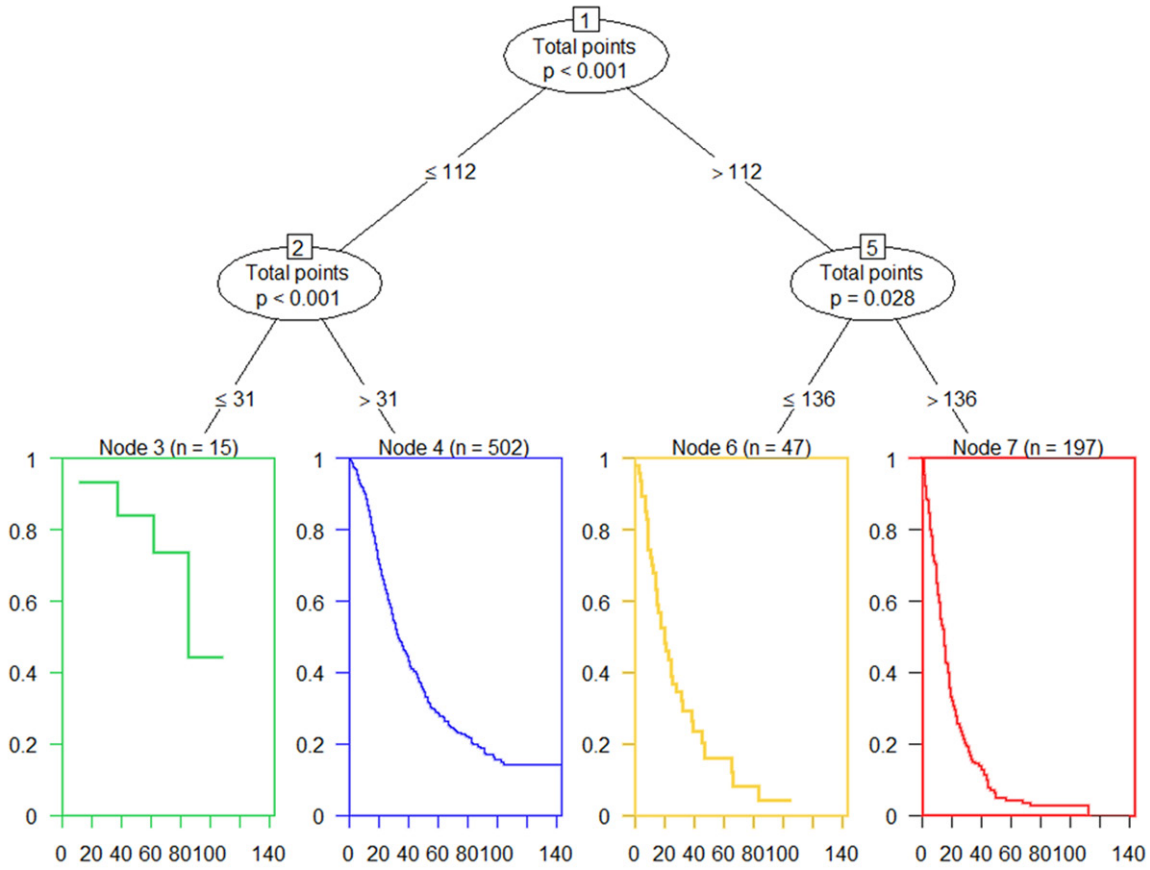
Supplementary Table 4. Discrimination measures and hazard ratios evaluated between two datasets

Measures of discrimination	NCKUH derivation cohort			CGMH validation cohort		
	Estimate	SE		Estimate	SE	
Harrell's C-index	0.643	0.012		0.636	0.013	
Gonen and Heller's K	0.617	0.010		0.617	0.010	
Royston & Sauerbrei's D	0.789	0.073		0.788	0.073	
Prognostic groups[a]	Hazard ratio	95% CI of HR	*p*-Value	Hazard ratio	95% CI of HR	*p*-Value
Low risk[b]	1			1		
Intermediate risk	3.384	1.399-8.183	0.007	3.405	1.269-9.136	0.015
High risk	6.318	2.486-16.054	< 0.001	4.712	1.697-13.087	0.003
Very high risk	9.676	3.968-23.594	< 0.001	8.366	3.107-22.526	< 0.001

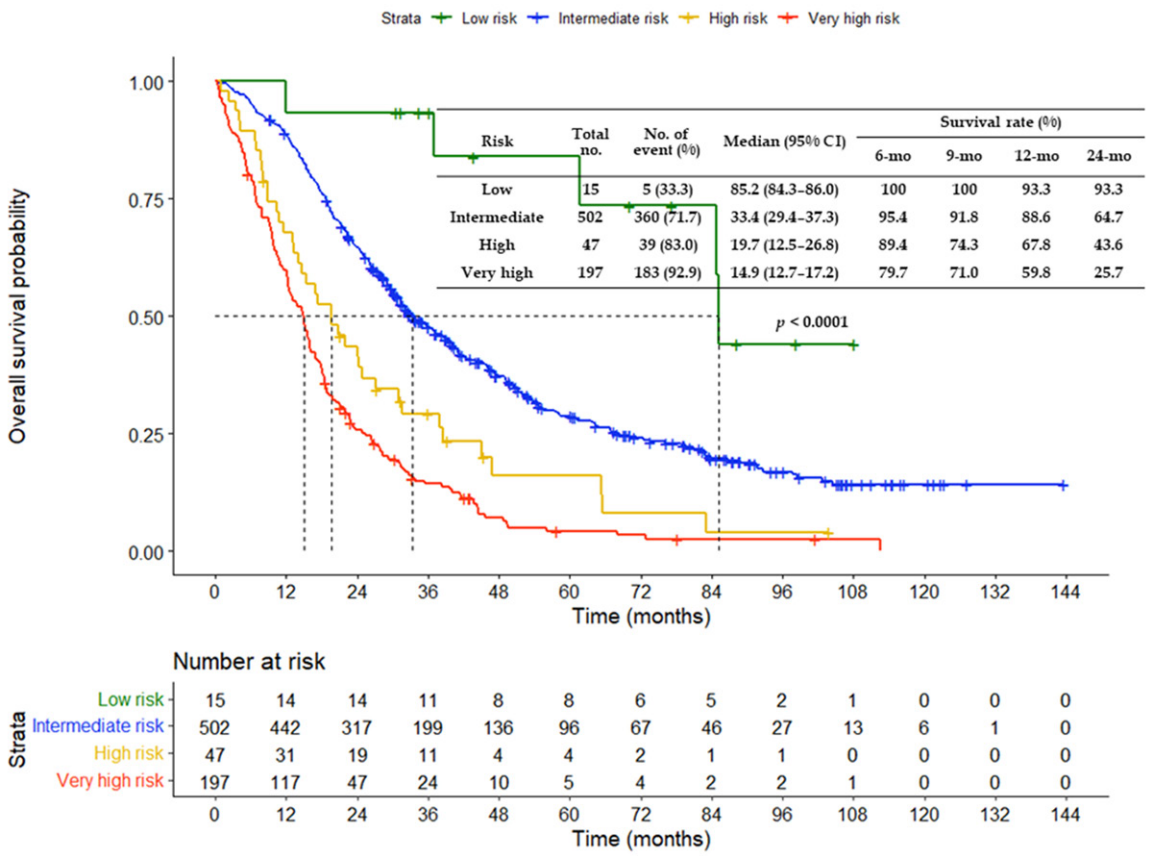
Abbreviations: CI, confidence interval; HR, hazard ratio; SE, standard error. [a]NCKUH Prognostic group was categorized using recursive partitioning analysis (RPA) to establish optimal cutoff points by Cox's model. [b]reference category.



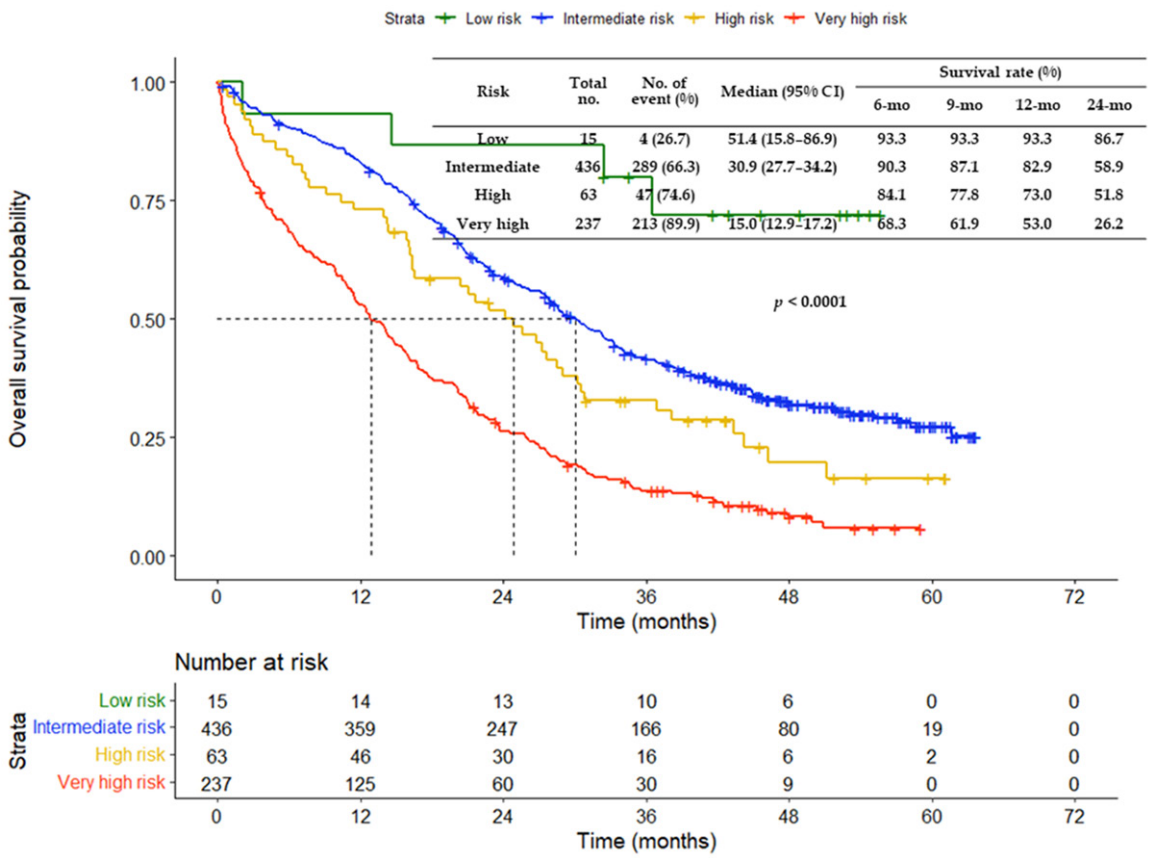
Supplementary Figure 2. Internal and external calibration curves. These show nomogram-predicted and actually observed OS. A. Internal calibration for 6-month OS. B. Internal calibration for 9-month OS. C. Internal calibration for 12-month OS. D. External calibration for 6-month OS. E. External calibration for 9-month OS. F. External calibration for 12-month OS. OS, overall survival.



Supplementary Figure 3. Survival analysis tree. A survival analysis tree was used to establish an optimal cutoff point to better predict overall survival.



Supplementary Figure 4. Kaplan-Meier curve of overall survival (OS) in NCKUH cohort. Kaplan-Meier curve for OS in the NCKUH derivation cohort according to risk groups. CI, confidence interval.



Risk	Total no.	No. of event (%)	Median (95% CI)	Survival rate (%)			
				6-mo	9-mo	12-mo	24-mo
Low	15	4 (26.7)	51.4 (15.8–86.9)	93.3	93.3	93.3	86.7
Intermediate	436	289 (66.3)	30.9 (27.7–34.2)	90.3	87.1	82.9	58.9
High	63	47 (74.6)		84.1	77.8	73.0	51.8
Very high	237	213 (89.9)	15.0 (12.9–17.2)	68.3	61.9	53.0	26.2

Supplementary Figure 5. Kaplan-Meier curve of overall survival (OS) in CGMH cohort. Kaplan-Meier curve for overall survival in the CGMH validation cohort according to risk groups. CI, confidence interval.