A nomogram prognostic model for clinical stage IA NSCLC: based on SEER databases

Bingchen Xu
Department of Thoracic Surgery, Sun Yat-sen University Cancer Center; State Key Laboratory of Oncology in South China; Collaborative Innovation Center for Cancer Medicine; Guangzhou

Ziming Ye
Department of Thoracic Surgery, Sun Yat-sen University Cancer Center; State Key Laboratory of Oncology in South China; Collaborative Innovation Center for Cancer Medicine; Guangzhou

Lianxin Zhu
Medical College of Nanchang University, Nanchang, China; Queen Mary University of London, London

Chunwei Xu
Department of Medical Oncology, Affiliated Jinling Hospital, Medical School of Nanjing University, Nanjing

Mingjian Lu
Department of Radiology, Affiliated Cancer Hospital and Institute of Guangzhou Medical University, Guangzhou, Guangdong510060, P. R

Qian Wang
Department of Respiratory Medicine, Affiliated Hospital of Nanjing University of Chinese Medicine, Jiangsu Province Hospital of Chinese Medicine, Nanjing

Wang Yao
Department of Interventional Oncology, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou

Zhihua Zhu (Zhuzhh@sysucc.org.cn)
Department of Thoracic Surgery, Sun Yat-sen University Cancer Center; State Key Laboratory of Oncology in South China; Collaborative Innovation Center for Cancer Medicine; Guangzhou

Research Article

Keywords: Nomogram, SEER, Prognosis, NSCLC, Treatment

Posted Date: June 22nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1757398/v1

Abstract

Background: The aim of this study was to establish and validate a nomogram model for accurate prediction of patients' survival with T1aN0M0 none small cell lung cancer (NSCLC).

Methods: The patients, diagnosed with the stage IA NSCLC from 2004-2015, were identified from the Surveillance, Epidemiology and End Results (SEER) database. The variables with a P value < 0.05 in a multivariate Cox regression were selected to establish the nomogram. The discriminative ability of the model was evaluated by the concordance index(C-index). The proximity of the nomogram prediction to the actual risk was depicted by a calibration plot. The clinical usefulness was estimated by the decision curve analysis (DCA). Survival curves were made with Kaplan-Meier method and compared by Log-Rank test.

Result: Eight variables, including treatment, age, sex, race, marriage, tumor size, histology and grade were selected to develop the nomogram model by univariate and multivariate cox regression. The C-index was 0.704 (95%CI, 0.694 to 0.714) in the training set and 0.713 (95%CI, 0.697 to 0.728) in the test set, which performed significantly better than 8^{th} edition AJCC TNM stage system (0.550, 95% CI, 0.408-0.683, P<0.001). The calibration curve showed that the prediction ability of 3-years and 5-years survival rate demonstrated a high degree of agreement between the nomogram model and the actual observation. The DCA curves also proved that the nomogram-assisted decisions could improve patient outcomes.

Conclusions: We established and validated a prognostic nomogram to predict 3-years and 5-years overall survival in stage IA NSCLC.

Introduction

Lung cancer remains the leading cause of cancer death with 1.8 million deaths each year, accounting for 18% of deaths worldwide (Hyuna et al 2021). The major histologic subtype is non-small-cell lung cancer (NSCLC) (Siegel et al2021). At present, early-stage NSCLC could be clinically diagnosed with ease due to the recent development of high resolution thin-slice computerized tomography (CT). However, due to the variety of treatments for early- stage NSCLC, the selection of suitable treatments for patient with different health conditions has always been a controversial problem.

Since the first prospective report of Lung Cancer Study Group (LCSG) declaring after randomized trial that limited resection has higher death rate and locoregional recurrence rate, the lobectomy with mediastinal lymph node dissection has become the standard treatment for T1N0M0 NSCLC (Ginsberg and Rubinstein 1995). In contrast, several studies reported that the survivals after sublobar resection or ablation were similar to those after lobectomy. The possible reason for the difference could be that the study of LCSG solely based on the tumor-node-metastasis (TNM) staging system, ignoring the influence of clinicopathologic features on patients' prognosis. Recently, a research in Japan which took tumor characteristics into consideration found that segmentectomy showed equal morbidity and mortality compared with lobectomy with diameter of tumor ≤ 2 cm and C/T ratio > 0.5 (Saji et al 2022). Similarly, in the JACS1303 including baseline factors, it was demonstrated that wedge resection might be equivalent to lobectomy or segmentectomy for selected patients older than 80-year (Mimae et al 2021). Therefore a new prognostic model is needed to include prognostic factors such as age, sex and tumor characteristics for clinical decision making.

Nomogram has been extensively used as a visualization prediction model. Many researches had confirmed that prediction model could improve the overall survival in patients with liver cancer, prostatic cancer and small cell lung cancer (Shan et al 2021; Smaletz et al 2002; Wang et al 2013). But to our knowledge, currently there is no similar

nomogram model for clinical decision making of early stage NSCLC. Therefore, in this research, we built an innovative nomogram model consisted of both TNM system and other prognostic factors in order to improve the personalized risk staging system and individual treatment decision making.

Methods

Patients

The Surveillance, Epidemiology and End Results (SEER) was our source of population-based data for nomogram development and validation. The version of our database was SEER*STAT 8.3.9.2. A flow-chat illustrating the methodology was used for extracting data of stage IA NSCLC in the SEER database during 2004–2015 (Fig. 1). Patients inclusion criteria were as follows: (I) diagnosed NSCLC between 2004 and 2015 by ICD-O-3 codes 8012, 8046, 8070–8072, 8140, 8250, 8255, 8260, 8480–8481, 8490, 8550, 8560, 8570; (II) clinically confirmed stage IA based on 8th edition of the TNM classification (Detterbeck et al 2016); (III) patients who underwent ablation, wedge resection, segmentectomy or lobectomy; (IV) patients with only one primary tumor. Exclusion criteria were as follows: (I) patients who received radiotherapy, adjuvant chemotherapy, neoadjuvant chemotherapy or other systemic treatment; (II) patients with severe complications or died within 30 days after surgery; (III) lost to follow-up; (IV) history of prior synchronous or metachronous malignancies. A total of 15,317 patients were extracted from the SEER database, among which 14,374 patients were selected into our cohort. Sample split function of catools in R was used to divide 14,374 samples in training cohort (N = 10,061) and validation cohort (N = 4,313) in a ratio of 7:3. The detail and code of the selection step were shown in the Table 1.

Table 1

Details patient selection criteria with variable names used and their effect on sample size.

Step	selection criteria	Code	Count
1	Select Respiratory Seer data with schema–lung	{Site and Morphology.Site recode ICD-O-3/WHO 2008} = ' Lung and Bronchus'	972941
2	Only patients with one primary	AND {Multiple Primary Fields.Sequence number} = 'One primary only'	695122
3	Year of diagnosis from 2004 to 2015	AND {Race, Sex, Year Dx, Registry, County.Year of diagnosis} ='2004','2005','2006','2007','2008','2009','2010','2011','2012','2013','2014','2015'	436736
4	SEER historic stage A = 1	AND {Stage - Summary/Historic.SEER historic stage A (1973–2015)} = 'Localized'	65092
5	Only patients with one malignant primary	AND {Multiple Primary Fields.First malignant primary indicator} = 'Yes'	65092
6	Select only patients with segmentectomy and wedge resection	AND {Therapy.RX Summ–Surg Prim Site (1998+)} = 12–13,15,21–22,33	26687
7	Select only IA stage(AJCC 6th) from 2005 to 2009 or IA stage (AJCC 7th) from 2004 to 2015	AND {Stage – 6th edition.Derived AJCC Stage Group, 6th ed (2004–2015)} = 'IA'	16970
8	Select Non-small cell lung cancer	AND {Site and Morphology.Histologic Type ICD-O-3} = 8012,8046,8070–8072,8140,8250,8255,8260,8480–8481,8490,8550,8560,8570	15317
9	Select 1mm < = tumor size < = 30mm	AND {Extent of Disease.CS tumor size (2004–2015)} = 1–30,991–993	15249
10	Exclude tracheal tumors	AND {Site and Morphology.Primary Site} != 340	15241
11	Exclude chemotherapy	AND {Therapy.Chemotherapy recode (yes, no/unk)} != 'Yes'	14798
12	Exclude radiotherapy	AND {Therapy.RX Summ–Surg/Rad Seq} = 'No radiation and/or cancer-directed surgery'	14516
13	Exclude survivaltime less than 1 month	AND {Cause of Death (COD) and Follow-up.Survival months} != 0	14374

Variables

In this study, we collected patient's baseline, treatment model and follow up information including age, sex, race, marital status, laterality, primary site, tumor size, histologic type, grade, marital status and survival time. The age should be converted to a categorial variable in this study with the cutoff value of 75-year determined by R software (Harrell 2010). Tumor stage was also an ordered variable needed to be converted to a categorial variable. Treatment methods consisted of lobectomy, segmentectomy, wedge resection and ablation, with each patient receiving only one treatment. The primary endpoint of the study was OS which was calculated from the date of diagnosed to patient's death or loss of follow-up.

Nomogram model development

The model was established using the training set and verified by the validation set in order to reduce over-fitting and upwardly biased estimates of performance. We used univariate Cox regression to select covariables associated with OS. Factors with a P-value < 0.10 were entered into the multivariate cox regression analysis. Then multivariate Cox regression was used to determine the independent prognostic factors with a P-value < 0.05. Based on the prognostic variables in the final model, the prediction model was established.

Validation of nomogram

The process of validation was used to obtain unbiased estimates of model performance and judge its applicability to different populations. Firstly, we chose bootstrapping, which was iteratively applied to randomly selected sample sets of the training cohort, to prevent overinterpretation. The discrimination ability of the prediction model was assessed by concordance index (C-index). Then we conducted calibration curves to evaluate the accuracy of the model by comparing the predicted survival time and observed survival rate of 3- and 5- years. Lastly, we would use decision curve analysis to assess whether nomogram-assisted decision could improve patient's outcomes.

Statistical analysis

Continuous variables were converted to categorical variables with the median used as a cutoff value, and categorical variables were denoted as percentages. The χ^2 test was applied to analyze the difference between two groups. Survival outcomes were evaluated by the Kaplan-Meier method and compared with the Log-Rank test. The univariate and multivariate cox proportional hazards regression was carried out to evaluate the strength of the association between the OS and the potential risk factors. All statistical analyses were conducted in the R software (version 4.1.0; http://www.r-project.org). R packages "gtsummary", "dplyr", "flextable", "survival", "catools", "rms", "crosstalk", "dynnom", "rsconnect", "cvauc", "regplot" were used.

Result

Baseline of patients

In this study, a total of 15,317 patients from the SEER database were included. The baseline characteristics of patients were shown in the Table 2. In comparison between the training set and the testing set, all the variables including surgery (P = 0.656), age (P = 0.692), sex (P = 0.667), race (P = 0.990), marriage (P = 0.724), laterality (P = 0.733), site (P = 0.237), size (P = 0.812), histology (P = 0.381), grade (P = 0.173) were not significantly different. The median survival times were 114 months (range 0 to 179 months) for the training dataset and 110 months for the testing dataset (range 0 to 179 months), respectively. The 3-year OS for the training and the testing datasets were 81.1% (95%CI 80.4–

81.9%) and 81.3% (95%CI 80.1–82.4%), respectively. The 5-year OS for the training and the testing datasets were 69.9% (95%CI 69.0-70.8%) and 70.0% (95%CI 68.6–71.4%), respectively.

Table 2
Characteristics of stage IA NSCLC patients from SEER Database.

	training set	test set	
	N = 10,061[1]	N = 4,313[1]	p-value[2]
surgery			0.656
ablation	110 (1.1%)	40 (0.9%)	
wedge resecton	2,156 (21%)	921 (21%)	
segmentectomy	583 (5.8%)	267 (6.2%)	
lobectomy	7,212 (72%)	3,085 (72%)	
Age			0.692
< 75 yr	7,389 (73%)	3,182 (74%)	
>=75 yr	2,672 (27%)	1,131 (26%)	
Sex			0.667
Female	5,828 (58%)	2,481 (58%)	
Male	4,233 (42%)	1,832 (42%)	
Race			0.990
White	8,519 (85%)	3,648 (85%)	
Black	805 (8.0%)	347 (8.0%)	
Other	737 (7.3%)	318 (7.4%)	
Marital status			0.724
Yes	5,602 (56%)	2,387 (55%)	
No	4,459 (44%)	1,926 (45%)	
Laterality			0.733
Lfet	4,103 (41%)	1,745 (40%)	
Right	5,958 (59%)	2,568 (60%)	
Primary Site			0.237
Upper lobe	6,397 (64%)	2,729 (63%)	
Middle lobe	434 (4.3%)	219 (5.1%)	
Lower lobe	3,139 (31%)	1,329 (31%)	

SCC squamous cell carcinoma, ADC Adenocarcinoma

[1]n (%)

[2]Pearson's Chi-squared test; Fisher's exact test

	training set	test set	
	N = 10,061[1]	N = 4,313[1]	p-value[2]
Other	91 (0.9%)	36 (0.8%)	
T stage			0.812
T1a	1,172 (12%)	513 (12%)	
T1b	5,314 (53%)	2,254 (52%)	
T1c	3,575 (36%)	1,546 (36%)	
Histologic type			0.381
SCC	2,398 (24%)	983 (23%)	
ADC	7,060 (70%)	3,074 (71%)	
Other	603 (6.0%)	256 (5.9%)	
Grade			0.173
I/II	6,829 (68%)	2,870 (67%)	
III/IV	2,604 (26%)	1,144 (27%)	
unknown	628 (6.2%)	299 (6.9%)	
SCC squamous cell carcinoma, ADC Adenocarcinoma			
[1]n (%)			
[2]Pearson's Chi-squared test; Fisher's exact test			

Selecting independent prognositic factors and establishing nomogram

The result of univariate analyses indicated that factors such as treatment (P < 0.001), age(P < 0.001), sex(P < 0.001), race(P < 0.001), marital statue(P < 0.001), tumor size(P < 0.001), histology(P < 0.001) and grade(P < 0.001) were associated with patient's prognosis (Table 3). The laterality and site of tumors were not the independent risk factors (P = 0.071 and P = 0.130). Further analysis in multivariable Cox regression demonstrated that the factors including treatment (P < 0.001), age (P < 0.001), sex (P < 0.001), race (P < 0.001), marital statue (P < 0.001), tumor size (P < 0.001), histology (P < 0.001) and grade (P < 0.001) were identified as independent prognostic factors (Table 3). The factors above were incorporated to develop the predicting model which was virtually presented in the form of a nomogram (Fig. 2). It could be observed that the treatment had the greatest impact on patient's survival, while grade or martial status showed moderate contributions to survival. Each straight line represented a factor with a corresponding number of points assigned to a particular magnitude of the variable. Then the point scores for all the variables were cumulated and located on the point scale for outcome. It was easy to get the final risk score and predict the OS at 3-,5-years for a specific patient.

Table 3
Univariate and multivariate Cox regression of OS in stage IA NSCLC patients.

	Univariate			Multivariate		
	HR[1]	95% CI[2]	p-value	HR	95% CI	p-value
Treatment						
Ablation	—	—		—	—	
Wedge resection	0.52	0.41, 0.64	< 0.001	0.67	0.54, 0.84	< 0.001
Segmentectomy	0.41	0.32, 0.52	< 0.001	0.57	0.44, 0.73	< 0.001
Lobectomy	0.30	0.24, 0.37	< 0.001	0.41	0.33, 0.51	< 0.001
Age						
< 75yr	—	—		—	—	
>=75yr	2.06	1.94, 2.20	< 0.001	1.82	1.70, 1.94	< 0.001
Sex						
Female	—	—		—	—	
Male	1.49	1.40, 1.58	< 0.001	1.50	1.41, 1.60	< 0.001
Race						
White	—	—		—	—	
Black	0.97	0.87, 1.09	0.600	0.97	0.86, 1.08	0.600
other	0.56	0.48, 0.65	< 0.001	0.63	0.55, 0.73	< 0.001
Marital status						
yes	—	—		—	—	
no	1.29	1.22, 1.38	< 0.001	1.27	1.19, 1.35	< 0.001
Laterality						
Left	—	—				
Right	0.95	0.89, 1.00	0.071			
Primary Site						
Upper lobe	—	—				
Middle lobe	0.89	0.76, 1.04	0.130			
Lower lobe	0.97	0.91, 1.04	0.400			
Other	1.01	0.74, 1.38	> 0.9			
T stage						
T1a	—	—		—	—	

[1]HR = Hazard Ratio, [2]CI = Confidence Interval.

	Univariate			Multivariate		
	HR[1]	95% CI[2]	p-value	HR	95% CI	p-value
T1b	1.14	1.03, 1.27	0.014	1.16	1.04, 1.29	0.006
T1c	1.51	1.35, 1.68	< 0.001	1.49	1.34, 1.67	< 0.001
Histologic type						
SCC	—	—		—	—	
ADC	0.53	0.49, 0.56	< 0.001	0.66	0.61, 0.70	< 0.001
other	1.00	0.89, 1.12	> 0.9	0.96	0.85, 1.08	0.500
Grade						
I/II	—	—		—	—	
III/IV	1.55	1.45, 1.65	< 0.001	1.26	1.18, 1.35	< 0.001
Unknown	1.24	1.10, 1.41	< 0.001	1.09	0.95, 1.24	0.200
[1]HR = Hazard Ratio, [2]CI = Confidence Interval.						

Calibration of the predict nomogram model

Bootstrapping was used with 1,000 repetitions resampling 3,000 samples a time to prevent data overinterpretation and obtain a relatively unbiased estimate of the model. The C-indexes of the nomogram were 0.704 (95%CI, 0.694–0.714) and 0.713 (95%CI, 0.697–0.728) in the training cohort and the test cohort, reflecting the adequacy to distinguish between patients with different treatment management. In contrast, the C-indexes for OS estimates of the eighth edition of the AJCC TNM staging system were 0.550 (95% CI, 0.408–0.683) in the training set and 0.548 (95%CI, 0.401–0.672) in the test set. The calibration curves for the model of 3- and 5-year OS were shown in the Fig. 3, demonstrating the proximity the nomogram prediction to the actual observation. Also, the result remained consistent in the test set. It was shown in the Fig. 4 that DCA exhibited great positive net benefits among all the threshold probabilities at different time points, indicating the favorable potential clinical effect of the predictive model.

Discussion

Stage I NSCLC is the very early stage of lung cancer. Due to its high heterogeneity and the development of technology of CT diagnosis, the treatment choices for different patients shown considerably large discrepancy (Donington et al 2017). With the continuous progress in precision and individualized treatment, the TNM staging system based solely on anatomical classification was not accurate enough for clinical usefulness. Therefore, it is urgent to develop a new prediction model to assist the surgeon in selection of suitable treatment strategies based on routinely clinicopathological variables. In this study, a predictive nomogram was established and verified to predict the prognosis of patients with stage IA NSCLC based on the pretreatment clinicopathological characteristics.

Many clinical factors could affect the survival of the patients of NSCLC (Merritt et al 2021; Koike et al 2013). Notably, the treatment management was an important prognostic factor for the lung cancer patients. Whitson et al. reported that lobectomy conferred superior overall (P < 0.0001) and cancer-specific (P = 0.005) 5-year survivals compared with segmentectomy in stage I adenocarcinoma (Whitson et al 2011). However, Dai et al. found that for T1aN0M0 NSCLC

patients not suitable for lobectomy, segmentectomy should be recommended to those with tumor size less than 2cm (Dai et al 2016). For patients who were not surgery candidates, thermal ablation demonstrated better results in overall survival and acceptable local control. Mimae et al. reported that the 3-years OS rates were slightly better after wedge resection than segmentectomy plus lobectomy for patients over 80-year (89.4%, 95% CI, 73.8–95.9% vs 75.8%, 95% CI, 62.0–85.2%; P = 0.14) (Mimae et al 2021). Similarly, Linda Willen et al. found that surgeries were more common in the younger age group and the usage of stereo-tactic body radiotherapy (SBRT) increased with the increase of age (< 69years 5.4%; >85years 35.8%) (Willen et al 2021). Gender had also been confirmed as an independent prognostic factor in surgically managed patients. Chansky et al, reported that female patients who underwent surgery alone had 5-year survival rates of 56.8%, which was better than their male counterparts with only 48.3% (P < 0.0001) (Chansky et al 2009). In conclusion, the selection of treatments should be made based on multiple factors after the comprehensive evaluation.

By incorporating these factors, our nomogram showed perfect discriminative ability. The C-indexes were 0.704 (95%CI, 0.694 to 0.714) in the training cohort and 0.713 in the test cohort (95%CI, 0.697 to 0.728), respectively. It was better than the C-indexes in TNM, which were 0.550 in the training set (95% CI, 0.408–0.683) and 0.548 in the test set (95%CI, 0.401–0.672, P < 0.001), respectively. The 3-year and 5-year validation curves also showed high degree of agreement with the actual situation. Besides, the DCA curves revealed favorable potential clinical usefulness. All the evaluation above confirmed that our nomogram was an excellent model with a powerful prognostic performance to predict the OS and assist the making of treatment decision.

Several nomograms had been established to predict the OS of NSCLC after surgery (Birim et al 2006; Zhang et al 2014; Liang et al 2015). However, there is no model developed to predict the treatment efficacy for previously untreated patients. Our model was designed to help doctors and patients choosing the best treatment. For individual, surgeon calculated scores according to patient's physical condition and tumor characteristics and then provided treatment recommendation based on the predicted survival. Our model compared the influences of different treatment strategies on the prognosis of NSCLC patients, which was easily practicable and in line with the actual situation of lung cancer.

Although some previous studies had demonstrated prognostic models for NSCLC, the results were slightly worse than ours. Yuan et al. showed a nomogram for cancer-specific survival of stage I NSCLC in 2019 with C-index of 0.64 (95%CI, 0.63–0.65) (Zeng et al 2019). The result of another study based on the lung adenocarcinoma in the stage I to III was only 0.69 (95%CI, 0.64–0.73) (Xie and Zhang 2021). The most important reason for the better result in this study was the large sample size. Our source of population-based data for model establishment and validation was the SEER program which currently captured 400,00 cancer cases annually and stored cancer data for approximately 34.6% of U.S. population from 18 (SEER) cancer registries. Our samples not only included different races and centers but also could be updated regularly, which guaranteed the timeliness of the model and minimized selection bias.

At present, many nomograms had been reported in the literature, but few were clinically applied. The possible reasons could be the difficulty in obtaining factors and high cost (Zeng et al 2019). David MJ et al (2012) reported a quantitative-PCR-baded assay to predict survival in resectable lung cancer. Some models were based on radiomics (Huang et al 2021) and artificial intelligence algorithm (Churchill et al 2021). These prediction models functioned well, but were too expensive with variables not easily available in all clinical settings. The variables in our model were easily available, which efficiently decreased the cost and increased the real-world practicality.

There were some limitations of this study. Firstly, this was a retrospective research and the data selection bias was unavoidable. Secondly, the cancer-specific survival (CSS) would be a more suitable metric compared to the OS, but determining the cause of death was unachievable in the SEER databases (Weiser et al 2011). Although we used

multivariable analysis to reduce the impact of confounding, there were still some unobtainable factors in SEER databases such as smoking, pulmonary function and gene mutation. As the SEER database was from the United States, more than 85% of our cases involved white people. Therefore, the prediction for the Asia-Pacific population needs further verification (Shi et al 2014).

Conclusion

We innovatively established a prognostic nomogram to predict the OS of patients diagnosed with stage IA NSCLC using clinicopathological factors, tumor characteristics and treatment modality with good discrimination and calibration ability. This model could be valuable in prognostic prediction and decision making of treatment.

Declarations

Funding Sources

This research was supported by Bethune Foundation. The grant number was zllcaxw-19. Zhihua Zhu has received research support.

Disclosures of Conflicts of Interest

All of author in this article disclosed no relevant relationships.

Author Contributions

Wang Yao and Zhihua Zhu wrote the review, edit and supervise the manuscript text. Ziming Ye and Mingjian Lu managed data curation. Chunwei Xu and Qian Wang administrated the project, resources and the investigation. Lianxin Zhu carried out the formal analysis and methodology. Bingchen Xu wrote the main manuscript text and validate the result. The manuscript was reviewed and approved by all authors for publication

References

1. Birim O, Kappetein AP, Waleboer M, et al. (2006) Long-term survival after non-small cell lung cancer surgery: development and validation of a prognostic model with a preoperative and postoperative mode. J Thorac Cardiovasc Surg;132(3):491–8. https://doi.org/10.1016/j.jtcvs.2006.04.010.
2. Chansky K, Sculier JP, Crowley JJ, Giroux D, Van Meerbeeck J, Goldstraw P. (2009) The International Association for the Study of Lung Cancer Staging Project: prognostic factors and pathologic TNM stage in surgically managed non-small cell lung cancer. J Thorac Oncol;4(7):792–801. https://doi.org/10.1097/JTO.0b013e3181a7716e.
3. Churchill IF, Gatti AA, Hylton DA, et al. (2021) An Artificial Intelligence Algorithm to Predict Nodal Metastasis in Lung Cancer. Ann Thorac Surg. https://doi.org/10.1016/j.athoracsur.2021.06.082.
4. Donington JS, Kim YT, Tong B, et al.(2018) Progress in the Management of Early-Stage Non-Small Cell Lung Cancer in 2017. J Thorac Oncol;13(6):767–778. https://doi.org/10.1016/j.jtho.2018.04.002.
5. Dai C, Shen J, Ren Y, et al. (2016) Choice of Surgical Procedure for Patients With Non-Small-Cell Lung Cancer ≤ 1 cm or > 1 to 2 cm Among Lobectomy, Segmentectomy, and Wedge Resection: A Population-Based Study. J Clin Oncol;34(26):3175–82. https://doi.org/10.1200/jco.2015.64.6729.

6. Ginsberg RJ, Rubinstein LV. (1995) Randomized trial of lobectomy versus limited resection for T1 N0 non-small cell lung cancer. Lung Cancer Study Group. Ann Thorac Surg;60(3):615–22; discussion 622-3. https://doi.org/10.1016/0003-4975(95)00537-u.
7. Huang L, Lin W, Xie D, et al. (2021) Development and validation of a preoperative CT-based radiomic nomogram to predict pathology invasiveness in patients with a solitary pulmonary nodule: a machine learning approach, multicenter, diagnostic study. Eur Radiol. https://doi.org/10.1007/s00330-021-08268-z.
8. Harrell FE (2010) Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis
9. Koike T, Koike T, Yoshiya K, Tsuchida M, Toyabe S. (2013) Risk factor analysis of locoregional recurrence after sublobar resection in patients with clinical stage IA non-small cell lung cancer. J Thorac Cardiovasc Surg;146(2):372–8. https://doi.org/10.1016/j.jtcvs.2013.02.057.
10. Kratz JR, He J, Van Den Eeden SK, et al. (2012) A practical molecular assay to predict survival in resected non-squamous, non-small-cell lung cancer: development and international validation studies. Lancet;379(9818):823–32. https://doi.org/10.1016/s0140-6736(11)61941-7.
11. Liang W, Zhang L, Jiang G, et al. (2015) Development and validation of a nomogram for predicting survival in patients with resected non-small-cell lung cancer. J Clin Oncol;33(8):861–9. https://doi.org/10.1200/jco.2014.56.6661.
12. Mimae T, Saji H, Nakamura H, et al.(2021) Survival of Octogenarians with Early-Stage Non-small Cell Lung Cancer is Comparable Between Wedge Resection and Lobectomy/Segmentectomy: JACS1303. Ann Surg Oncol;28(12):7219–7227. https://doi.org/10.1245/s10434-021-09835-w.
13. Merritt RE, Abdel-Rasoul M, Fitzgerald M, D'Souza DM, Kneuertz PJ. (2021) Nomograms for Predicting Overall and Recurrence-free Survival From Pathologic Stage IA and IB Lung Cancer After Lobectomy. Clin Lung Cancer;22(4):e574-e583.. https://doi.org/10.1016/j.cllc.2020.10.009.
14. Sung H, Ferlay J, Siegel RL, et al.(2021) Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin;71(3):209–249. https://doi.org/10.3322/caac.21660.
15. Siegel RL, Miller KD, Fuchs HE, Jemal A. (2021) Cancer Statistics, 2021. CA Cancer J Clin;71(1):7–33.. https://doi.org/10.3322/caac.21654.
16. Saji H, Okada M, Tsuboi M, et al. (2022) Segmentectomy versus lobectomy in small-sized peripheral non-small-cell lung cancer (JCOG0802/WJOG4607L): a multicentre, open-label, phase 3, randomised, controlled, non-inferiority trial. The Lancet;399(10335):1607–1617. https://doi.org/10.1016/s0140-6736(21)02333-3.
17. Shan Q, Shi J, Wang X, et al. (2021) A new nomogram and risk classification system for predicting survival in small cell lung cancer patients diagnosed with brain metastasis: a large population-based study. BMC Cancer;21(1):640. https://doi.org/10.1186/s12885-021-08384-5.
18. Smaletz O, Scher HI, Small EJ, et al. (2002) Nomogram for overall survival of patients with progressive metastatic prostate cancer after castration. J Clin Oncol;20(19):3972–82. https://doi.org/10.1200/jco.2002.11.021.
19. Shi Y, Au JS, Thongprasert S, et al. (2014) A prospective, molecular epidemiology study of EGFR mutations in Asian patients with advanced non-small-cell lung cancer of adenocarcinoma histology (PIONEER). J Thorac Oncol;9(2):154–62. https://doi.org/10.1097/jto.0000000000000033.
20. Wang Y, Li J, Xia Y, et al. (2013) Prognostic nomogram for intrahepatic cholangiocarcinoma after partial hepatectomy. J Clin Oncol;31(9):1188–95. https://doi.org/10.1200/jco.2012.41.5984.

21. Whitson BA, Groth SS, Andrade RS, Maddaus MA, Habermann EB, D'Cunha J. (2011) Survival after lobectomy versus segmentectomy for stage I non-small cell lung cancer: a population-based analysis. Ann Thorac Surg;92(6):1943–50. https://doi.org/10.1016/j.athoracsur.2011.05.091.
22. Willen L, Berglund A, Bergstrom S, et al. (2021) Are older patients with non-small cell lung cancer receiving optimal care? A population-based study. Acta Oncol:1–9. https://doi.org/10.1080/0284186X.2021.2000637.
23. Weiser MR, Gonen M, Chou JF, Kattan MW, Schrag D. (2011) Predicting survival after curative colectomy for cancer: individualizing colon cancer staging. J Clin Oncol;29(36):4796–802. https://doi.org/10.1200/JCO.2011.36.5080.
24. Xie H, Zhang JF, Li Q. (2021) Development of a Prognostic Nomogram for Patients with Lung Adenocarcinoma in the Stages I, II, and III Based on Immune Scores. Int J Gen Med;14:8677–8688. https://doi.org/10.2147/ijgm.S337934.
25. Zhang Y, Sun Y, Xiang J, Zhang Y, Hu H, Chen H. (2014) A clinicopathologic prediction model for postoperative recurrence in stage Ia non-small cell lung cancer. J Thorac Cardiovasc Surg;148(4):1193–9. https://doi.org/10.1016/j.jtcvs.2014.02.064.
26. Zeng Y, Mayne N, Yang CJ, et al. (2019) A Nomogram for Predicting Cancer-Specific Survival of TNM 8th Edition Stage I Non-small-cell Lung Cancer. Ann Surg Oncol;26(7):2053–2062. https://doi.org/10.1245/s10434-019-07318-7.

Figures

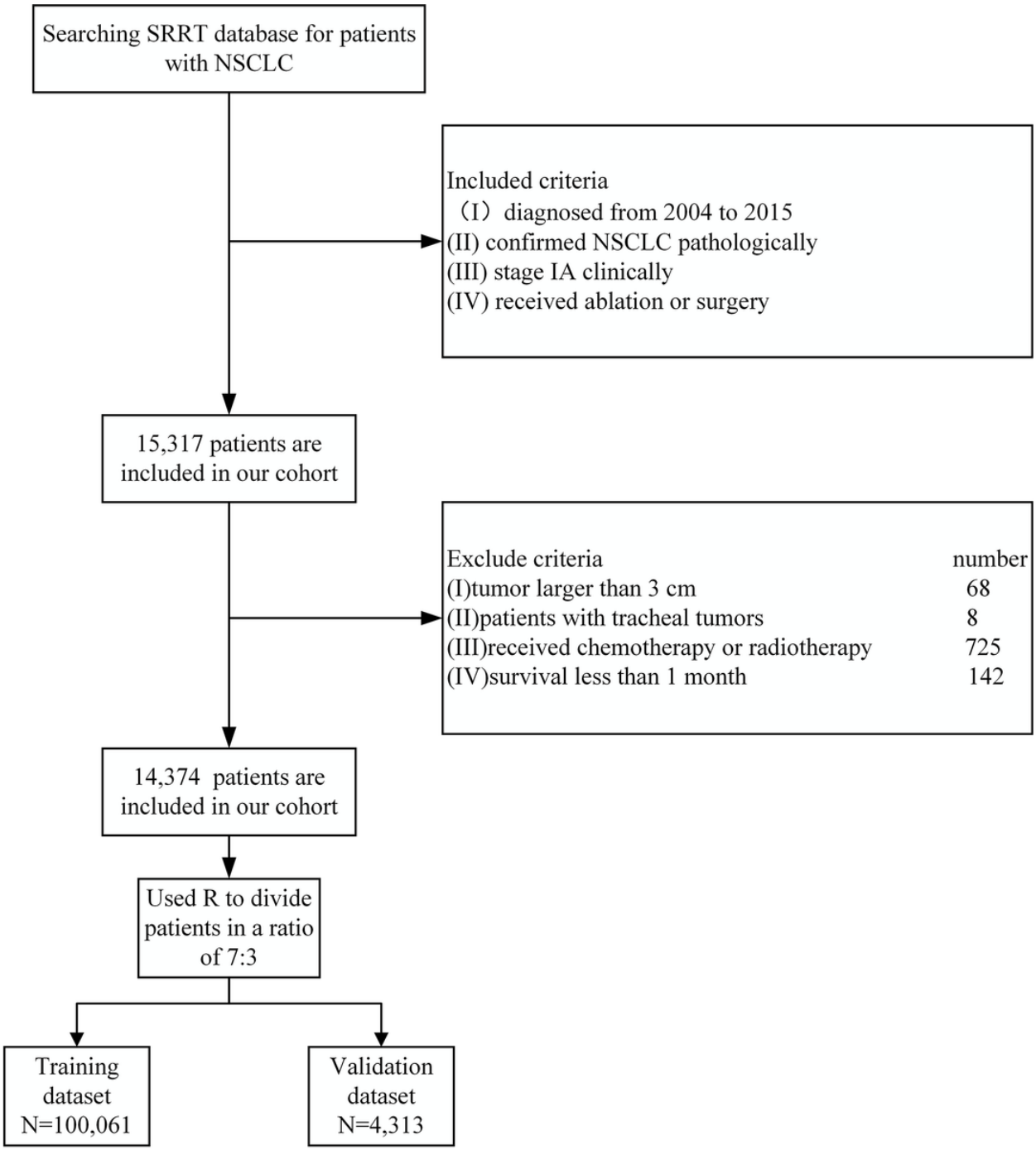

Figure 1

Flow-chart illustrating the steps to extract the case of NSCLC from SEER

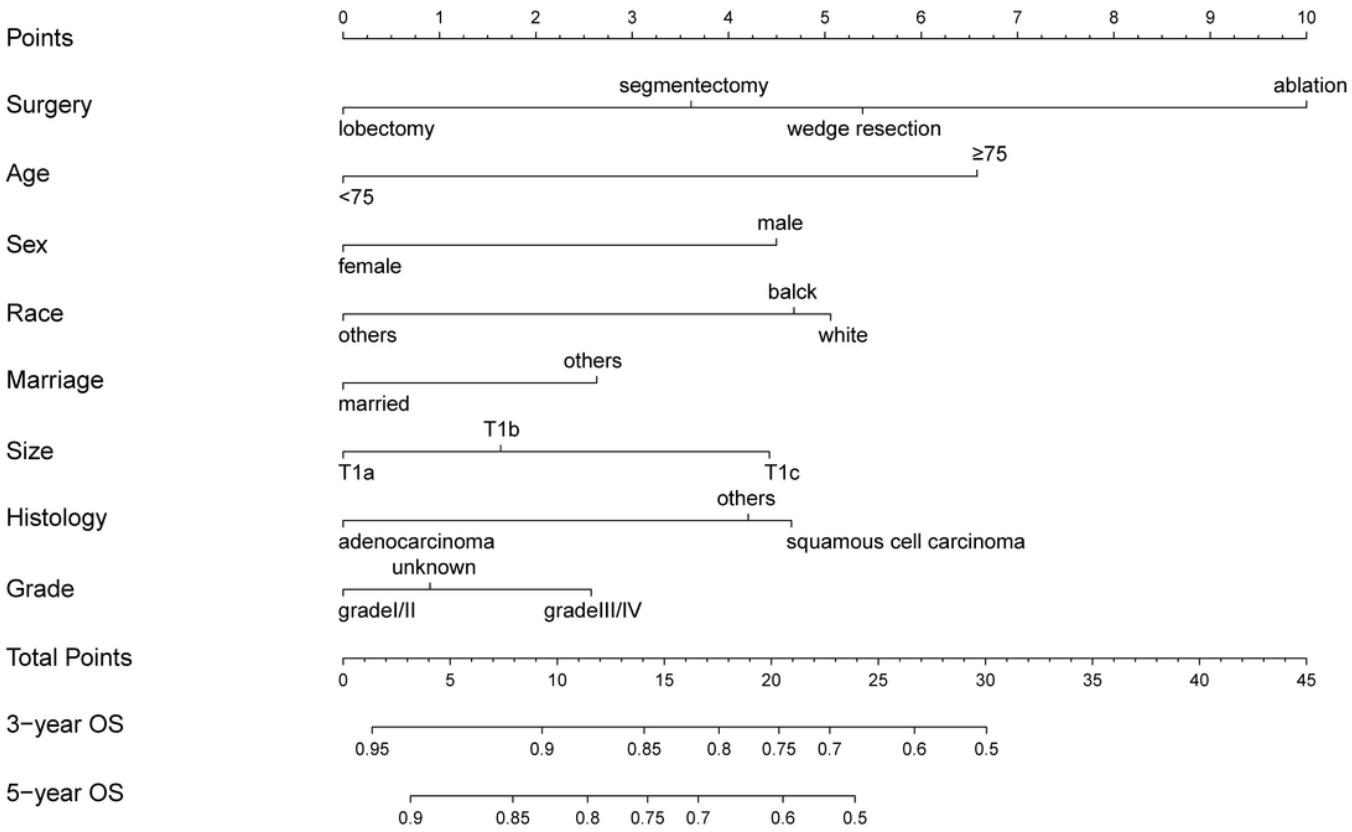

Figure 2

3 years- and 5 years- overall survival nomogram for patients with stage IA NSCLC

Figure 3

Calibration plots of nomogram model. (a)3 years of training set; (b)5 years of training set; (c)3 years of test set; (d)5 years of test set; Gray line, actual observation; Red line, prediction survival rate.

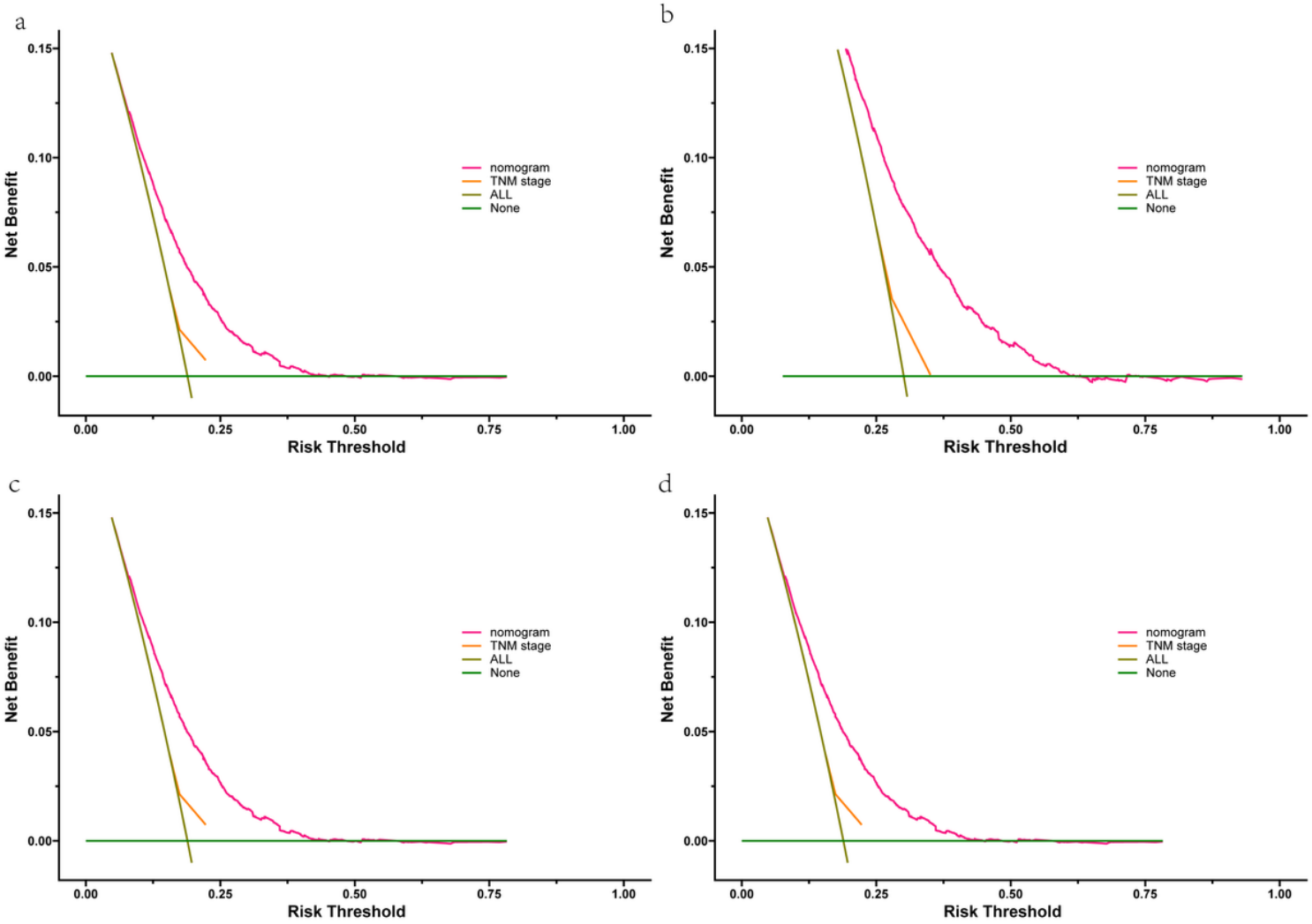

Figure 4

Decision curve analysis (DCA) for the nomogram model and the 8th edition AJCC TNM staging system. (a) Comparison of 3 years of DCA in training set; (b) Comparison of 5 years of DCA in training set; (c) Comparison of 3 years of DCA in test set; (d)Comparison of 5 years of DCA in test set. The horizontal line (Risk Threshold) referred to the motality and the vertical line referred to the net benefit. When the death rate exceeded a certain value, the clinician should start treatment for patients. After start the treatment, the patients who acturally will die would be benefited from the treatment (positive predictivity), and those who will not die would be harmed by the treatment (false positive rate). The net benefit = positive predictivity – false positive rate. The green horizontal line represents none of the patients received intervention and the other oblique green line represent all patients have been treated.