Am J Cancer Res 2024;14(11):5154-5177
www.ajcr.us /ISSN:2156-6976/ajcr0158713

Original Article

Machine learning-based prognostic models and factors influencing the benefit of surgery on primary lesion for patients with lung cancer brain metastases

Xixi Zhao[1*], Chaofan Li[2*], Mengjie Liu[2], Zeyao Feng[2], Xinyu Wei[2], Yusheng Wang[3], Jiaqi Zhao[4], Shuqun Zhang[2], Jingkun Qu[2]

[1]Department of Radiation Oncology, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi, P. R. China; [2]Department of Surgical Oncology, The Comprehensive Breast Care Center, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi, P. R. China; [3]Department of Otolaryngology, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi, P. R. China; [4]Department of Cardiology, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi, P. R. China. [*]Equal contributors.

Received June 24, 2024; Accepted October 25, 2024; Epub November 15, 2024; Published November 30, 2024

Abstract: Brain metastasis is very common in lung cancer and it's a fatal disease with extremely poor prognosis. Until now, there has been a lack of accurate and efficient prognostic models for patients with lung cancer brain metastases (LCBM), and the factors influencing the effectiveness of the surgery on primary lesion for these patients remain unclear. We used 7 machine learning algorithms to create prognostic models to predict the overall survival (OS) of LCBM based on the data from the Surveillance Epidemiology and End Results. Then, a series of validation methods, including area under the curve values, receiver operating characteristic curve analysis, calibration curves, decision curve analysis and external data validation were used to confirm the high discrimination, accuracy, and clinical applicability of the XGBoost models. Propensity score matching adjusted analysis was conducted for further stratified analysis to find factors influencing the benefit of surgery on primary lesion for LCBM. Models using XGBoost algorithm performed best. Surgery on primary lesion was a favorable independent prognostic factor for LCBM. Age > 70 years old, blacks, grade IV, stage T4, N3, other distant organ metastases, squamous cell carcinoma, large cell carcinoma and no radiation were all unfavorable factors of primary lung tumor surgery for the prognosis of LCBM. Our study is the first one to create highly accurate AI models to predict the OS of LCBM. Our in-depth stratified analysis found some influence factors of surgery on primary lesion for the prognosis of LCBM.

Keywords: Lung cancer, brain metastasis, XGBoost, surgery, SEER

Introduction

Nowadays, lung cancer (LC) is the top 1 most prevalent of all cancers and it is also the leading cause of cancer-related deaths, accounting for 20% of all cancer deaths worldwide [1-3]. The extremely high rate of distant metastases in LC is the main cause of poor prognosis; non-small cell lung cancer (NSCLC) accounts for 85% of all LC patients [4] and around 40% of these patients have distant metastases at initial diagnosis [5], while about 70% of patients with small cell lung cancer (SCLC) have distant metastases at primary diagnosis [6]. Among all distant metastases in LC, central nervous system metastases are the most common and devastating, with 10-20% having such lesions at the time of diagnosis and up to 50% developing brain metastases (BM) as the disease progresses [6, 7], and 40-50% of all BM are from LC [8, 9].

Patients with lung cancer brain metastases (LCBM) generally have a particularly poor prognosis, with an average median survival time of only 3-6 months [9-12], posing remarkable clinical challenges. Meanwhile, BM often cause epilepsy, cognitive impairment or other neurological dysfunction, which further substantially reduce the quality of life [7, 13]. Therefore, the main concern of the patients with LCBM and their families is how long they will live, and we

https://doi.org/10.62347/PRFQ9244

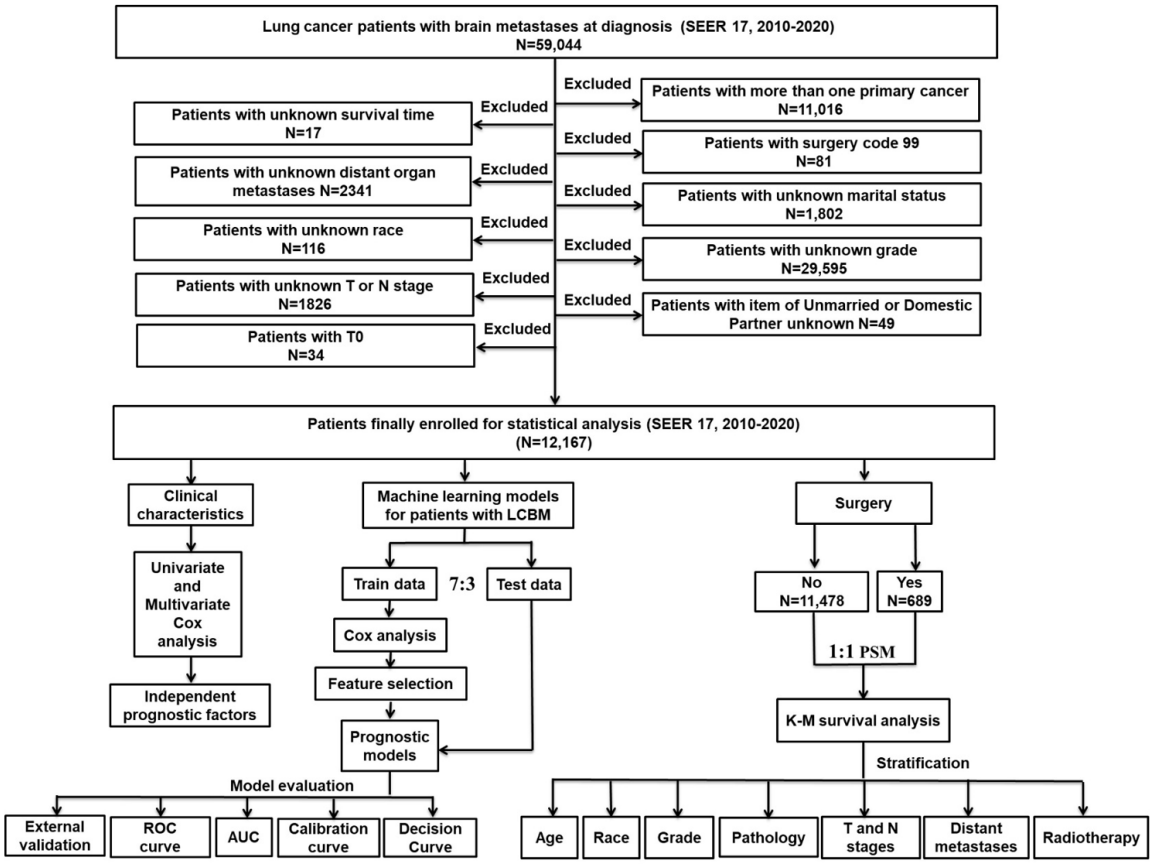


Figure 1. The flowchart detailed the procedure for carrying out the study and analysis of data. SEER: Surveillance Epidemiology and End Results, ROC: Receiver operating characteristic, AUC: the area under the curve, LCBM: lung cancer brain metastases, PSM: propensity score matching.

need an accurate prognostic model to answer this question and help optimise patient management.

Former studies have built several prognostic models by nomogram to predict the prognosis of patients with LCBM, however, the accuracy of these nomograms is nowhere near enough (C-index or AUC value < 0.7) [14-17]. In addition, some nomograms can only predict the prognosis of patients with SCLC [14, 15, 18], and other nomograms can only work in LC patients with simple BM [17], or predict the lung cancer specific survival (LCSS) of BM [16]. Therefore, we lack a model that can be applied widely and accurately to predict overall survival (OS) of LCBM. To achieve this, machine learning can help us create artificial intelligence (AI) predictive models and make them more accurate in large amounts of high-dimensional and multimodal data [19-21]. We used seven types of machine learning algorithms to create prognostic models and found that Extreme Gradient Boosting (XGBoost), one of the numerous machine learning algorithms [19, 20], was the most accurate.

Moreover, we unexpectedly found that a very small percentage of patients with de novo LCBM in the Surveillance Epidemiology and End Results (SEER) database had a primary lung tumor surgically removed, which was not usually recommended. The results even showed that resection of the primary lung tumor was an independent favorable prognostic factor for LCBM. We therefore further conducted stratified analyses to investigate the factors affecting therapeutic effect of the primary lung tumor surgery in patients with LCBM.

Our study examines the prognostic factors in patients with LCBM using the most recent SEER database and is the first one to create highly accurate AI models to predict the 6-month, 1-

Table 1. Baseline characteristics of patients with LCBM in overall, training and test sets

Characteristic	level	Overall (n=12167)	Training set (n=8516)	Test set (n=3651)	*P*
		Cases (%)	Cases (%)	Cases (%)	
Age	50	773 (6.35)	535 (6.28)	238 (6.52)	0.867
	50-59	2907 (23.89)	2020 (23.72)	887 (24.29)	
	60-69	4427 (36.39)	3123 (36.67)	1304 (35.72)	
	70-79	3056 (25.12)	2134 (25.06)	922 (25.25)	
	80+	1004 (8.25)	704 (8.27)	300 (8.22)	
Sex	Female	5716 (46.98)	4028 (47.30)	1688 (46.23)	0.29
	Male	6451 (53.02)	4488 (52.70)	1963 (53.77)	
Race	White	9393 (77.20)	6576 (77.22)	2817 (77.16)	0.384
	Black	1478 (12.15)	1017 (11.94)	461 (12.63)	
	Other	1296 (10.65)	923 (10.84)	373 (10.22)	
Marriage status	Married	6597 (54.22)	4592 (53.92)	2005 (54.92)	0.579
	Divorced/Separated	1778 (14.61)	1266 (14.87)	512 (14.02)	
	Single	2298 (18.89)	1604 (18.84)	694 (19.01)	
	Widowed	1494 (12.28)	1054 (12.38)	440 (12.05)	
Months from diagnosis to therapy	0	4804 (39.48)	3395 (39.87)	1409 (38.59)	0.383
	≥ 1	5832 (47.93)	4063 (47.71)	1769 (48.45)	
	Unknown	1531 (12.58)	1058 (12.42)	473 (12.96)	
Pathological types	Adenocarcinoma	7126 (58.57)	5016 (58.90)	2110 (57.79)	0.547
	Squamous cell carcinoma	1641 (13.49)	1143 (13.42)	498 (13.64)	
	Small cell carcinoma	1189 (9.77)	808 (9.49)	381 (10.44)	
	Large cell carcinoma	257 (2.11)	178 (2.09)	79 (2.16)	
	Others	1954 (16.06)	1371 (16.10)	583 (15.97)	
Grade	Grade I	414 (3.40)	287 (3.37)	127 (3.48)	0.897
	Grade II	2737 (22.50)	1922 (22.57)	815 (22.32)	
	Grade III	8034 (66.03)	5629 (66.10)	2405 (65.87)	
	Grade IV	982 (8.07)	678 (7.96)	304 (8.33)	
T stage	T1	1226 (10.08)	856 (10.05)	370 (10.13)	0.765
	T2	3432 (28.21)	2384 (27.99)	1048 (28.70)	
	T3	3052 (25.08)	2157 (25.33)	895 (24.51)	
	T4	4457 (36.63)	3119 (36.63)	1338 (36.65)	
N stage	N0	2845 (23.38)	1984 (23.30)	861 (23.58)	0.841
	N1	1182 (9.71)	821 (9.64)	361 (9.89)	
	N2	5694 (46.80)	4008 (47.06)	1686 (46.18)	
	N3	2446 (20.10)	1703 (20.00)	743 (20.35)	
Chemotherapy	No/Unknown	5246 (43.12)	3654 (42.91)	1592 (43.60)	0.489
	Yes	6921 (56.88)	4862 (57.09)	2059 (56.40)	
Radiation	None/Unknown	2929 (24.07)	2010 (23.60)	919 (25.17)	0.067
	Yes	9238 (75.93)	6506 (76.40)	2732 (74.83)	
Surgery	No	11478 (94.34)	8037 (94.38)	3441 (94.25)	0.814
	Yes	689 (5.66)	479 (5.62)	210 (5.75)	
Bone	No	8315 (68.34)	5834 (68.51)	2481 (67.95)	0.563
	Yes	3852 (31.66)	2682 (31.49)	1170 (32.05)	
Liver	No	10190 (83.75)	7134 (83.77)	3056 (83.70)	0.946
	Yes	1977 (16.25)	1382 (16.23)	595 (16.30)	
Lung	No	9088 (74.69)	6352 (74.59)	2736 (74.94)	0.701
	Yes	3079 (25.31)	2164 (25.41)	915 (25.06)	

LCBM, lung cancer brain metastases.

Table 2. Univariate and multivariate Cox analysis of LCBM characteristics

	Univariate Cox analysis						Multivariate Cox analysis					
	OS			LCSS			OS			LCSS		
	HR	95% CI	*P* Value	HR	95% CI	*P* Value	HR	95% CI	*P* Value	HR	95% CI	*P* Value
Age												
< 50	Reference						Reference			Reference		
50-59	1.32	1.21-1.44	***	1.34	1.22-1.46	***	1.21	1.11-1.33	***	1.23	1.12-1.36	***
60-69	1.48	1.36-1.60	***	1.48	1.35-1.61	***	1.32	1.20-1.44	***	1.32	1.21-1.45	***
70-79	1.89	1.74-2.06	***	1.90	1.73-2.08	***	1.59	1.45-1.74	***	1.59	1.45-1.76	***
≥ 80	2.38	2.15-2.63	***	2.35	2.12-2.62	***	1.68	1.50-1.89	***	1.66	1.47-1.87	***
Sex												
Female	Reference						Reference			Reference		
Male	1.27	1.22-1.32	***	1.27	1.22-1.32	***	1.23	1.18-1.28	***	1.23	1.17-1.28	***
Race												
White	Reference			Reference			Reference			Reference		
Black	1.03	0.97-1.09	0.35	1.02	0.96-1.08	0.56	0.96	0.90-1.02	0.19	0.95	0.89-1.02	0.15
Others	0.67	0.63-0.71	***	0.67	0.62-0.71	***	0.70	0.66-0.76	***	0.70	0.65-0.76	***
Marriage status												
Married	Reference			Reference			Reference			Reference		
Divorced/Separated	1.25	1.19-1.32	***	1.25	1.18-1.33	***	1.24	1.17-1.32	***	1.25	1.17-1.33	***
Single	1.15	1.10-1.21	***	1.13	1.07-1.20	***	1.06	1.00-1.12	0.06	1.05	0.99-1.11	0.13
Widowed	1.34	1.26-1.42	***	1.34	1.26-1.42	***	1.14	1.07-1.22	***	1.15	1.07-1.24	***
Months from diagnosis to therapy												
0 month	Reference			Reference			Reference			Reference		
≥ 1 month	0.81	0.78-0.84	***	0.81	0.78-0.85	***	0.78	0.75-0.82	***	0.78	0.75-0.82	***
Pathological type												
Adenocarcinoma	Reference						Reference			Reference		
Squamous cell	1.67	1.57-1.76	***	1.67	1.58-1.78	***	1.46	1.37-1.55	***	1.49	1.40-1.59	***
Small cell	1.36	1.28-1.45	***	1.37	1.28-1.47	***	1.32	1.21-1.44	***	1.34	1.22-1.47	***
Large cell	1.33	1.17-1.51	***	1.34	1.17-1.53	***	1.32	1.14-1.52	***	1.36	1.17-1.58	***
Others	1.47	1.40-1.55	***	1.42	1.34-1.50	***	1.32	1.24-1.40	***	1.29	1.21-1.38	***
Grade												
Grade I	Reference			Reference			Reference			Reference		
Grade II	1.10	0.98-1.23	0.09	1.06	0.94-1.19	0.33	1.16	1.03-1.31	*	1.11	0.98-1.25	0.10
Grade III	1.46	1.31-1.63	***	1.40	1.25-1.57	***	1.36	1.21-1.52	***	1.30	1.16-1.47	***
Grade IV	1.57	1.39-1.78	***	1.51	1.33-1.72	***	1.37	1.19-1.59	***	1.29	1.11-1.50	***

T Stage												
T1	Reference			Reference			Reference			Reference		
T2	1.26	1.17-1.35	***	1.28	1.19-1.38	***	1.29	1.20-1.40	***	1.30	1.20-1.41	***
T3	1.47	1.37-1.58	***	1.51	1.40-1.63	***	1.43	1.33-1.55	***	1.45	1.33-1.57	***
T4	1.54	1.44-1.65	***	1.59	1.48-1.71	***	1.45	1.34-1.57	***	1.46	1.35-1.59	***
N Stage												
N0	Reference			Reference			Reference			Reference		
N1	1.06	0.98-1.14	0.14	1.06	0.98-1.15	0.12	1.06	0.98-1.14	0.18	1.07	0.98-1.16	0.13
N2	1.20	1.14-1.26	***	1.21	1.15-1.27	***	1.17	1.11-1.24	***	1.18	1.11-1.25	***
N3	1.22	1.15-1.29	***	1.23	1.16-1.31	***	1.22	1.14-1.30	***	1.23	1.14-1.31	***
Chemotherapy												
No/unknown	Reference			Reference			Reference			Reference		
Yes	0.36	0.34-0.37	***	0.37	0.36-0.39	***	0.39	0.37-0.40	***	0.40	0.38-0.42	***
Radiotherapy												
No/unknown	Reference			Reference			Reference			Reference		
Yes	0.59	0.56-0.61	***	0.62	0.59-0.65	***	0.92	0.86-0.97	**	0.94	0.89-1.00	0.05
Surgery												
No	Reference			Reference			Reference			Reference		
Yes	0.47	0.43-0.52	***	0.47	0.43-0.51	***	0.55	0.50-0.60	***	0.54	0.49-0.60	***
Bone metastasis												
No	Reference						Reference			Reference		
Yes	1.16	1.11-1.21	***	1.19	1.14-1.24	***	1.19	1.13-1.24	***	1.21	1.15-1.27	***
Liver metastasis												
No	Reference			Reference			Reference			Reference		
Yes	1.48	1.41-1.55	***	1.49	1.41-1.57	***	1.38	1.30-1.46	***	1.37	1.29-1.46	***
Lung metastasis												
No	Reference			Reference			Reference			Reference		
Yes	1.22	1.17-1.28	***	1.24	1.19-1.30	***	1.12	1.06-1.18	***	1.14	1.08-1.20	***

*P < 0.05, **P < 0.01, ***P < 0.001. LCBM, lung cancer brain metastases.

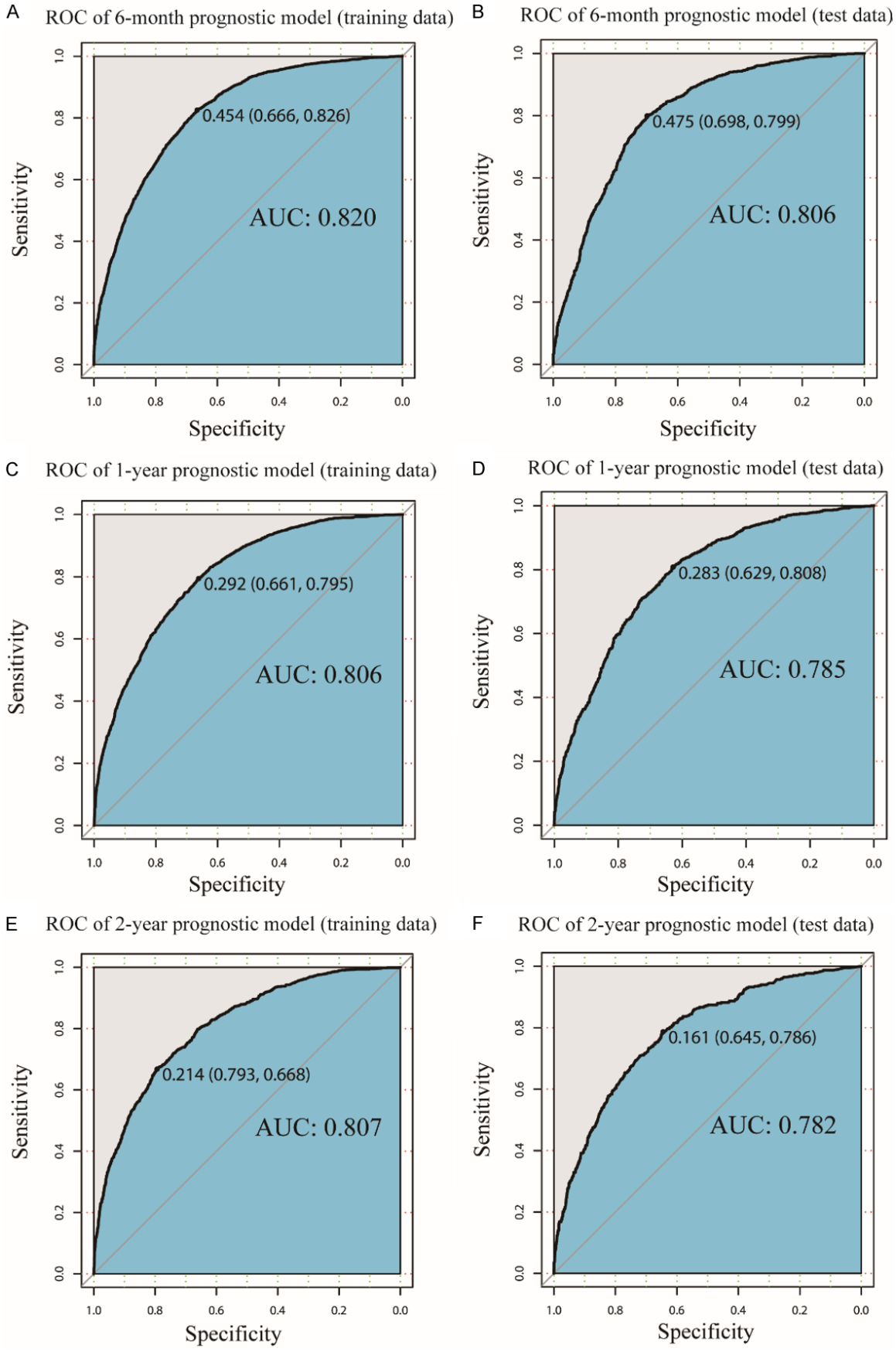

A
ROC of 6-month prognostic model (training data)
0.454 (0.666, 0.826)
AUC: 0.820
B
ROC of 6-month prognostic model (test data)
0.475 (0.698, 0.799)
AUC: 0.806
C
ROC of 1-year prognostic model (training data)
0.292 (0.661, 0.795)
AUC: 0.806
D
ROC of 1-year prognostic model (test data)
0.283 (0.629, 0.808)
AUC: 0.785
E
ROC of 2-year prognostic model (training data)
0.214 (0.793, 0.668)
AUC: 0.807
F
ROC of 2-year prognostic model (test data)
0.161 (0.645, 0.786)
AUC: 0.782
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0

Figure 2. ROC curve of the XGBoost models. ROC curve for the (A) 6-month prognostic model in training data; (B) 6-month prognostic model in test data; (C) 1-year prognostic model in training data; (D) 1-year prognostic model in test data; (E) 2-year prognostic model in training data; (F) 2-year prognostic model in test data. ROC: Receiver operating characteristic, AUC: the area under the curve.

Table 3. The AUC value of prognostic models constructed by machine learning algorithms on test data

	6-month survival	1-year survival	2-year survival
XGBoost	0.806	0.785	0.782
LR	0.777	0.761	0.765
ANN	0.775	0.753	0.751
RSF	0.757	0.746	0.729
DT	0.695	0.716	0.511
KNN	0.647	0.611	0.627
SVM	0.714	0.593	0.526

AUC, the area under the curve; LR, logistic regression; XGBoost, extreme gradient boosting; ANN, approximate nearest neighbor; RSF, random survival forest; DT, decision tree; KNN, K-Nearest Neighbor; SVM, support vector machine.

and 2-year OS in patients with LCBM. Additionally, we further investigated the role of primary lung tumor surgery in patients with LCBM, and found that over 70 years old, black people, squamous cell and large cell LC, Grade IV, T4, N3, no radiation therapy and other distant metastases are the unfavorable prognostic factor for primary lung tumor surgery in patients with LCBM, which have never been reported before. This work provides insight into patients with LCBM and is useful for prognostic forecast and clinical management.

Materials and methods

Data source and study design

The workflow for this study design and analysis is illustrated in **Figure 1**. Data were obtained from the SEER database for patients with LCBM in this study [SEER research data, 17 Regs, (changes 2010-2020); version 8.4.1]. Inclusion criteria were as follows: 1) only LC; 2) brain metastases; 3) age ≥ 18 years; 4) all cancer patients showed evidence of the histopathological and morphological from the International Classification of Cancer Diseases, Third Revision (ICD-O-3). Exclusion criteria were as follows: 1) patients with more than one primary cancer; 2) patients whose survival time is unknown; 3) patients with surgery code 99; 4) sit of distant organ metastases is unknown; 5) marital status is unknown; 6) race is unknown; 7) grade is unknown; 8) T or N stage is unknown; 9) the item of unmarried or domestic partner unknown. Follow-up until patient death, loss to follow-up or December 31, 2020.

Construction and validation of machine learning models

Patients with LCBM were sorted into two sets: a training set (n=8516) and a test set (n=3651) at random in a 7:3 ratio. Statistically significant characteristics by multivariable Cox regression analyses in the training set were included in our machine learning models, including age, sex, race, marital status, time from diagnosis to treatment, pathological pattern, grade, T and N stage, therapy method of chemotherapy, radiation and primary lung tumor surgery, distant organ metastasis of bone, liver, and Lung, to predict 6-month, 1- and 2-year overall survival of patients with LCBM. A response variable for survival information was obtained prior to the initiation of the training programme, with 1 indicating survival and 0 indicating death. We used 7 machine learning algorithms, such as logistic regression (LR), extreme gradient boosting (XGBoost), approximate nearest neighbor (ANN), random survival forest (RSF), decision tree (DT), K-Nearest Neighbor (KNN) and support vector machine (SVM) to create models and compared the area under the curve (AUC) values of them, and found that XGBoost models performed best. The main parameters of the XGBoost models were shown in Supplementary Table 1 and how to select these hyperparameters for the XGBoost model was illustrated in Supplementary Figure 1. Receiver operating characteristic curve (ROC), calibration curve and decision curve analysis (DCA) were used to confirm the high discrimination, accuracy, and clinical applicability of the XGBoost models.

External validation data

To further validate the XGBoost models, 32 patients' data diagnosed with LCBM between

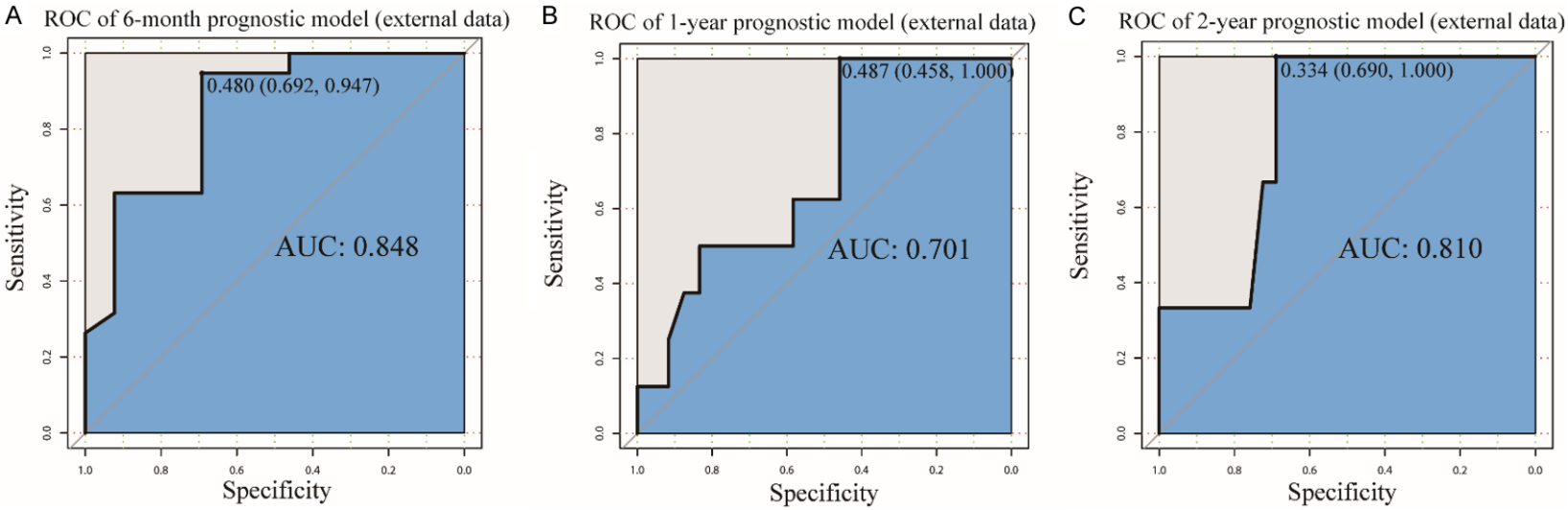


Figure 3. ROC curve of external validation data set. ROC curve for the (A) 6-month prognostic model; (B) 1-year prognostic model; (C) 2-year prognostic model in external validation data set. ROC: Receiver operating characteristic, AUC: the area under the curve.

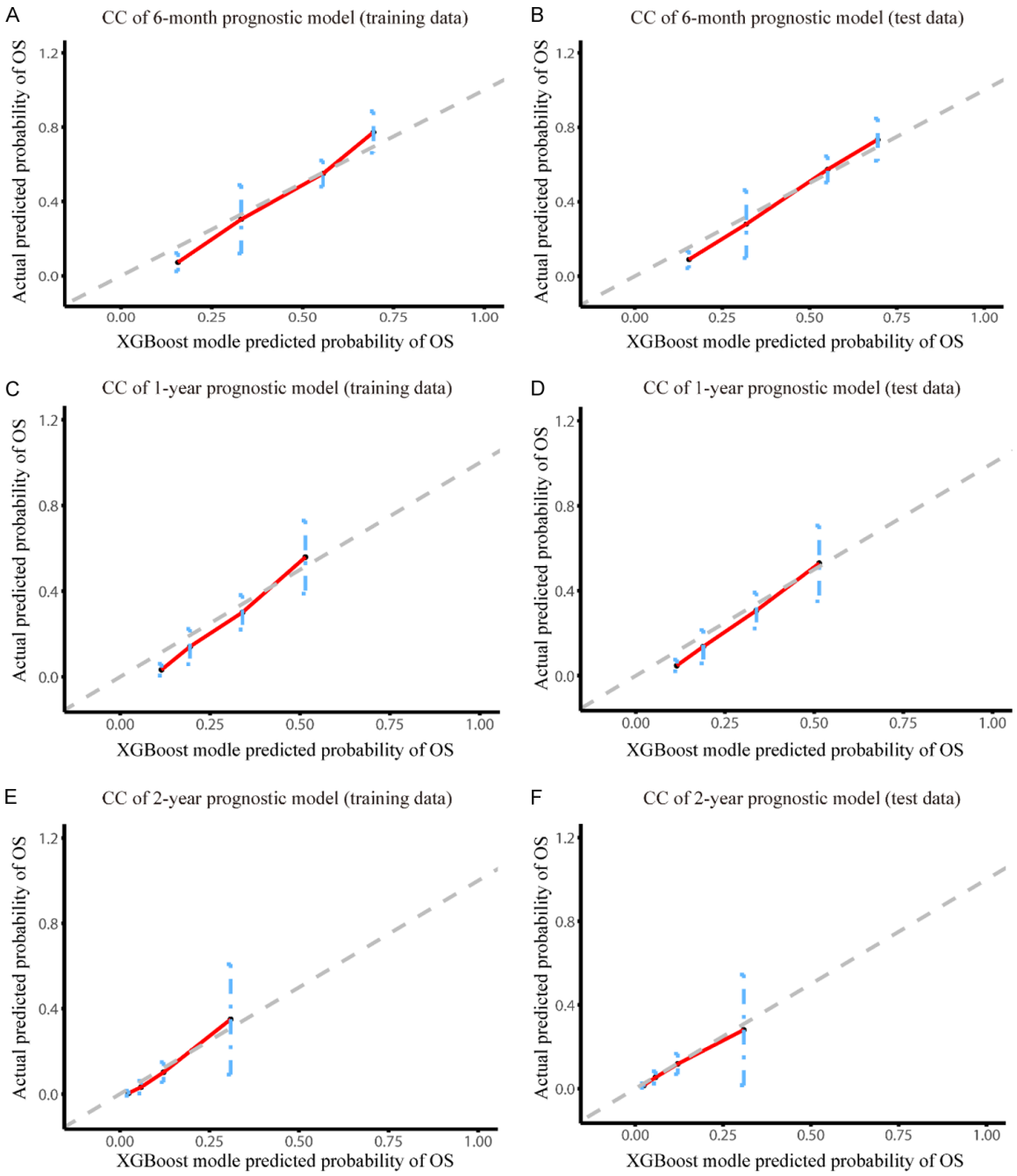


Figure 4. Calibration curve of the XGBoost models. Calibration curve for the (A) 6-month prognostic model in training data; (B) 6-month prognostic model in test data; (C) 1-year prognostic model in training data; (D) 1-year prognostic model in test data; (E) 2-year prognostic model in training data; (F) 2-year prognostic model in test data. CC: Calibration curve.

December 2018 and December 2021 was collected in the Second Affiliated Hospital of Xi'an Jiaotong University. The exclusion criteria of patient selection were as the same as it for SEER data.

Statistical analysis

Univariate and multivariate Cox regression analyses were applied to evaluate the mortality risk and independent prognostic factors.

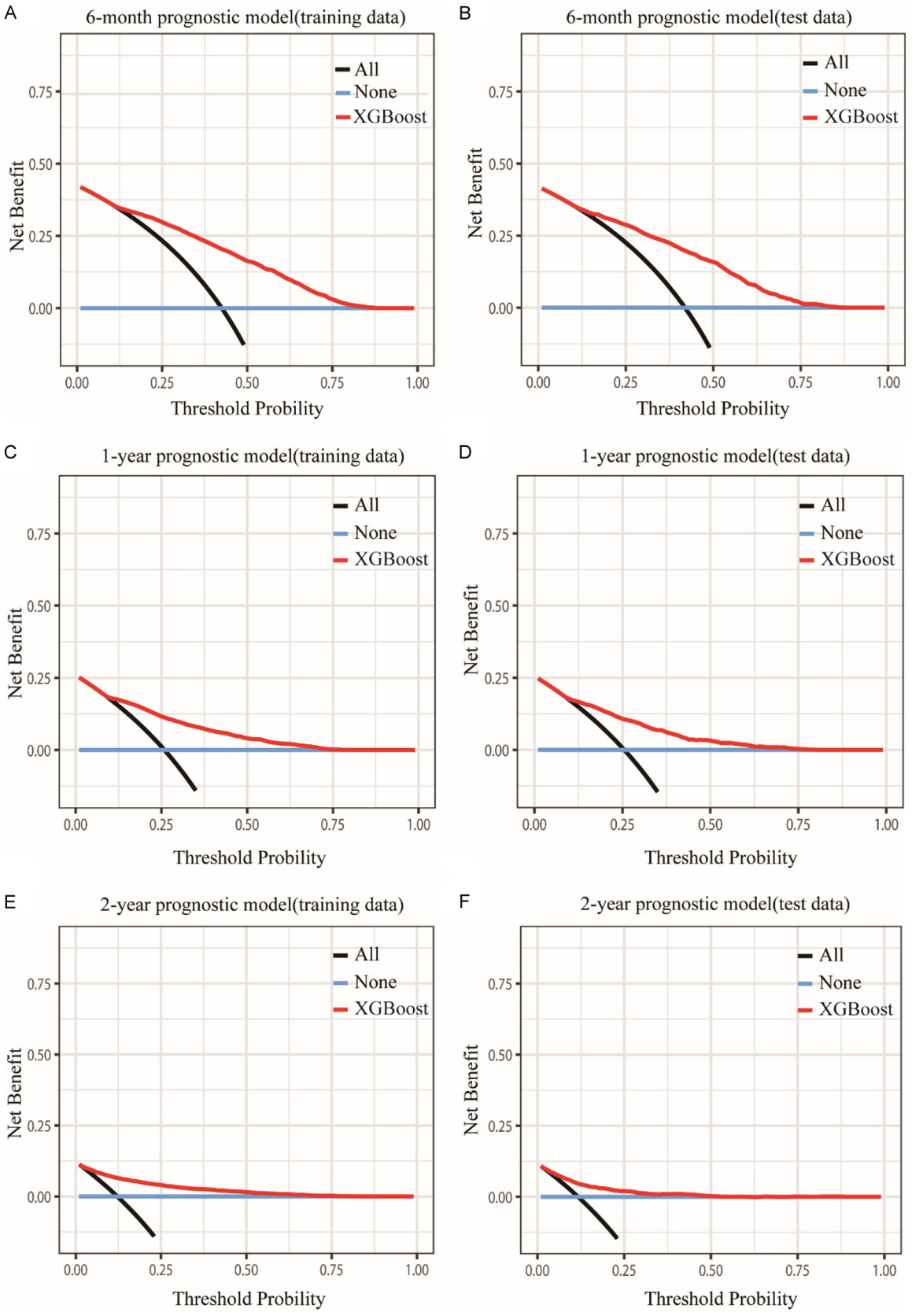


Figure 5. Decision curve analysis of the XGBoost models. Decision curve for the (A) 6-month prognostic model in training data; (B) 6-month prognostic model in test data; (C) 1-year prognostic model in training data; (D) 1-year prognostic model in test data; (E) 2-year prognostic model in training data; (F) 2-year prognostic model in test data.

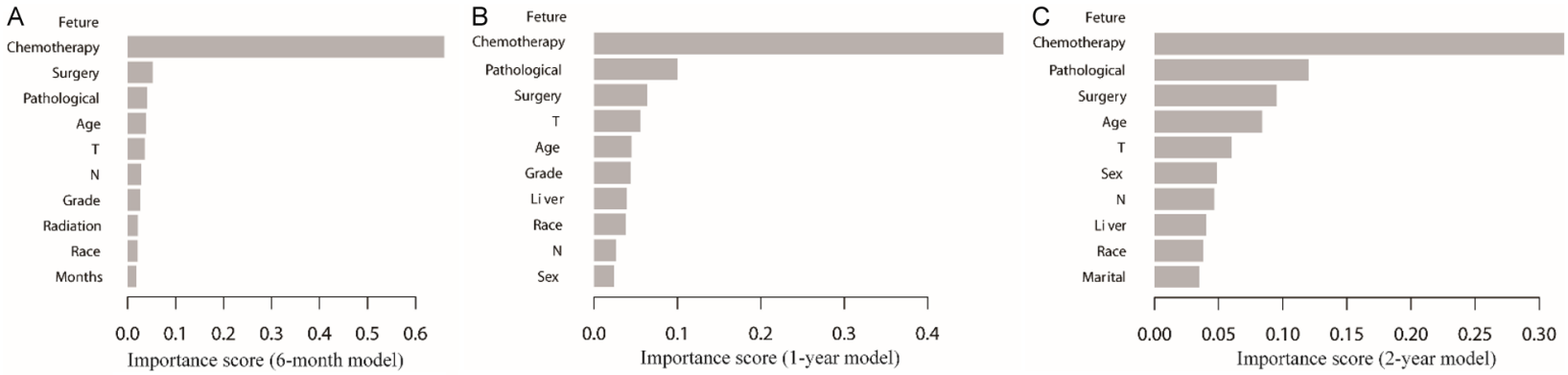


Figure 6. The important ranking of clinical characteristics in the XGBoost prognostic models. The important ranking of clinical characteristics (A) in the 6-month prognostic model; (B) in the 1-year prognostic model; (C) in the 2-year prognostic model.

Table 4. Comparison of patient features by surgery before and after propensity score matching (PSM)

Characteristics	Unmatched Cohort			1:1 propensity score matched (PSM) Cohort		
	Surgery not given	Surgery	Unadjusted	Surgery not given	Surgery	PSM-adjusted
	N=11478 (%)	N=689 (%)	*P* value	N=685 (%)	N=685 (%)	*P* value
Age			***			0.87
< 50	711 (6.19)	62 (9.00)		55 (8.03)	61 (8.91)	
50-59	2710 (23.61)	197 (28.59)		201 (29.34)	195 (28.47)	
60-69	4160 (36.24)	267 (38.75)		279 (40.73)	266 (38.83)	
70-79	2916 (25.41)	140 (20.32)		128 (18.69)	140 (20.44)	
≥ 80	981 (8.55)	23 (3.34)		22 (3.21)	23 (3.36)	
Sex			0.83			0.59
Female	5389 (46.95)	327 (47.46)		337 (49.20)	326 (47.59)	
Male	6089 (53.05)	362 (52.54)		348 (50.80)	359 (52.41)	
Race			**			0.78
White	8824 (76.88)	569 (82.58)		561 (81.90)	565 (82.48)	
Black	1412 (12.30)	66 (9.58)		63 (9.20)	66 (9.64)	
Others	1242 (10.82)	54 (7.84)		61 (8.91)	54 (7.88)	
Marriage status			**			0.72
Married	6185 (53.89)	412 (59.80)		413 (60.29)	409 (59.71)	
Divorced/Separated	1687 (14.70)	91 (13.21)		78 (11.39)	91 (13.28)	
Single	2168 (18.89)	130 (18.87)		132 (19.27)	129 (18.83)	
Widowed	1438 (12.53)	56 (8.13)		62 (9.05)	56 (8.18)	
Months from diagnosis to therapy			***			0.96
0 month	4417 (38.48)	387 (56.17)		382 (55.77)	384 (56.06)	
≥ 1 month	5530 (48.18)	302 (43.83)		303 (44.23)	301 (43.94)	
Pathological type			***			0.36
Adenocarcinoma	6707 (58.43)	419 (60.81)		437 (63.80)	416 (60.73)	
Squamous cell	1553 (13.53)	88 (12.77)		73 (10.66)	88 (12.85)	
Small cell	1164 (10.14)	25 (3.63)		33 (4.82)	25 (3.65)	
Large cell	227 (1.98)	30 (4.35)		22 (3.21)	30 (4.38)	
Others	1827 (15.92)	127 (18.43)		120 (17.52)	126 (18.39)	
Grade			***			0.46
Grade I	388 (3.38)	26 (3.77)		35 (5.11)	25 (3.65)	
Grade II	2522 (21.97)	215 (31.20)		204 (29.78)	214 (31.24)	
Grade III	7622 (66.41)	412 (59.80)		404 (58.98)	411 (60.00)	
Grade IV	946 (8.24)	36 (5.22)		42 (6.13)	35 (5.11)	
T Stage			***			0.84
T1	1099 (9.57)	127 (18.43)		111 (16.20)	123 (17.96)	
T2	3162 (27.55)	270 (39.19)		279 (40.73)	270 (39.42)	
T3	2895 (25.22)	157 (22.79)		156 (22.77)	157 (22.92)	
T4	4322 (37.65)	135 (19.59)		139 (20.29)	135 (19.71)	
N Stage			***			0.71
N0	2515 (21.91)	330 (47.90)		326 (47.59)	326 (47.59)	
N1	1068 (9.30)	114 (16.55)		112 (16.35)	114 (16.64)	
N2	5491 (47.84)	203 (29.46)		214 (31.24)	203 (29.64)	
N3	2404 (20.94)	42 (6.10)		33 (4.82)	42 (6.13)	
Chemotherapy			**			0.09
No/unknown	4983 (43.41)	263 (38.17)		230 (33.58)	261 (38.10)	
Yes	6495 (56.59)	426 (61.83)		455 (66.42)	424 (61.90)	
Radiotherapy			*			0.55
No/unknown	2787 (24.28)	142 (20.61)		149 (21.75)	139 (20.29)	
Yes	8691 (75.72)	547 (79.39)		536 (78.25)	546 (79.71)	

Bone metastasis			***			1.00
No	7718 (67.24)	597 (86.65)		592 (86.42)	593 (86.57)	
Yes	3760 (32.76)	92 (13.35)		93 (13.58)	92 (13.43)	
Liver metastasis			***			0.45
No	9537 (83.09)	653 (94.78)		656 (95.77)	649 (94.74)	
Yes	1941 (16.91)	36 (5.22)		29 (4.23)	36 (5.26)	
Lung metastasis			***			0.78
No	8465 (73.75)	623 (90.42)		623 (90.95)	619 (90.36)	
Yes	3013 (26.25)	66 (9.58)		62 (9.05)	66 (9.64)	

*P < 0.05, **P < 0.01, ***P < 0.001. PSM, propensity score matching.

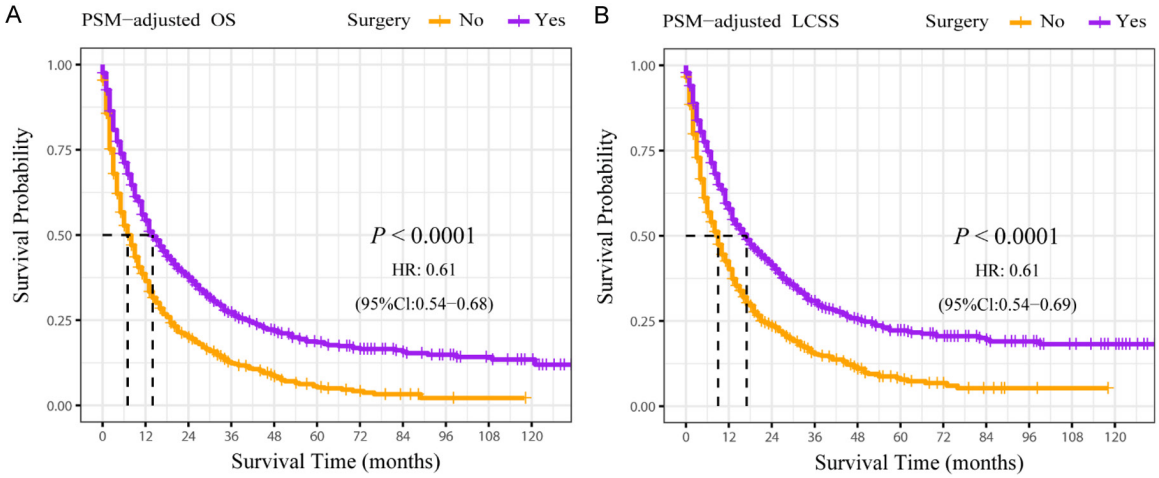


Figure 7. OS and LCSS of LCBM patients underwent primary lung tumor surgery or not after PSM adjustment. PSM adjusted Kaplan-Meier (K-M) survival analysis: (A) OS of LCBM underwent primary lung tumor surgery or not; (B) LCSS of LCBM underwent primary lung tumor surgery or not. PSM: propensity score matching, OS: overall survival, LCSS: lung cancer specific survival, LCBM: lung cancer brain metastases.

Among the patients with LCBM who receiving primary lung tumor surgery and those who did not were paired on a 1:1 basis by propensity score matching (PSM), using the variables identified in the univariate Cox regression analyses. This approach helps to reduce confounding and makes it easier to compare the effects of treatment in different subgroups [22]. Kaplan-Meier (K-M) curve analysis was then performed [23] within the PSM-adjusted population, stratified by all the independent prognostic factors of patients with LCBM. In this study, all statistical analyses were performed using R software (version 4.1.3). Statistical significance was defined as P < 0.05.

Ethics statement

Ethical review and approval were waived for this study due to the fact that the data are fully de-identified and no intervention on patients was performed.

Results

Clinical characteristics of patients with LCBM

Overall, data from 12167 patients with LCBM were extracted from the SEER database (2010-2020). Their clinicopathological features are summarized in **Table 1** and detailed below. These patients aged 60-69 accounted for the largest proportion (36.36%) of the total number of patients, while patients aged under 50 accounted for the smallest proportion, only 6.35%. The percentage of male patients (53.02%) is slightly higher than the percentage of female patients (46.98%). White patients are the main population, accounting for 77.2% of the total population, and 12.15% were of black ethnicity. In terms of marital status, 54.2% of the patients were married, while 18.89% were single. Approximately 39.48% of the patients began therapy immediately after diagnosis, while 47.93% started therapy > 1 month follow-

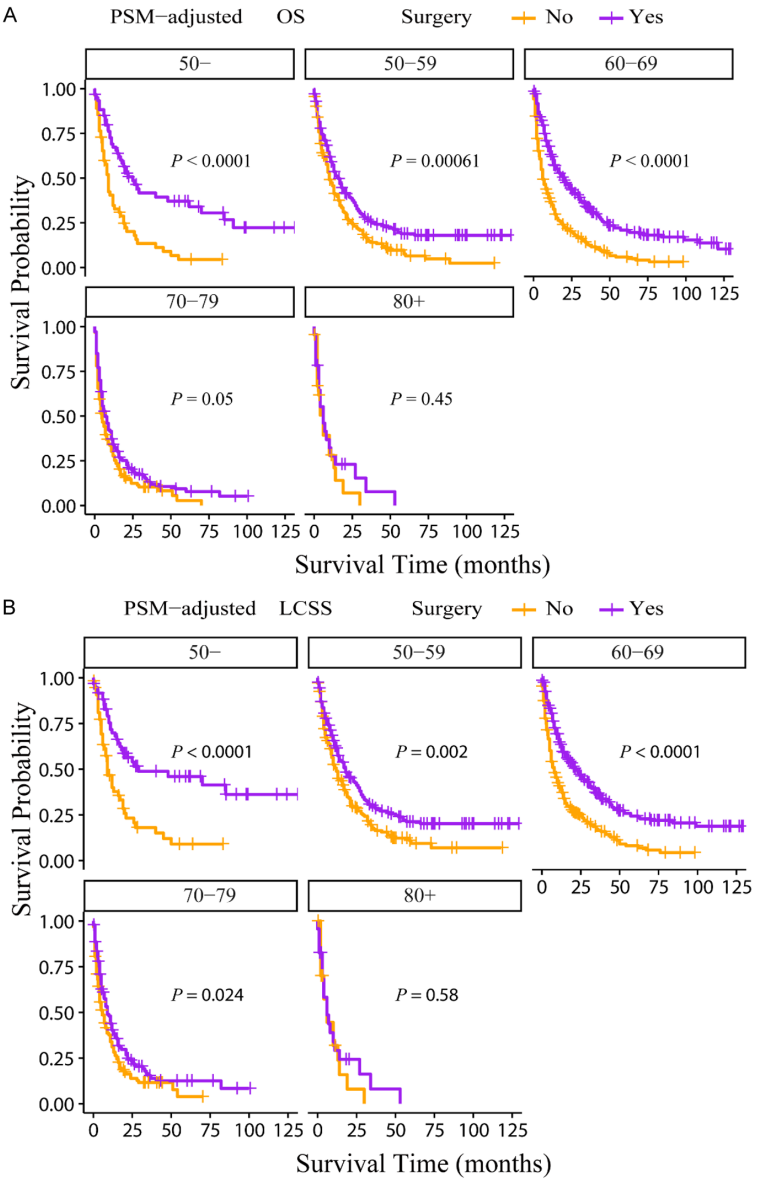


Figure 8. OS and LCSS of LCBM patients underwent primary lung tumor surgery or not after PSM adjustment (Stratified by age). A: OS of LCBM patients with different age groups; B: LCSS of LCBM patients with different age groups. PSM: propensity score matching, OS: overall survival, LCSS: lung cancer specific survival, LCBM: lung cancer brain metastases.

ing diagnosis. Among the pathological types, adenocarcinoma accounted for 58.57%, squamous cell carcinoma for 13.49%, small cell lung cancer for 9.77% and large cell carcinoma for only 2.11%. Grade III was the most common (66.03%), while Grade I was the least prevalent (3.40%). Staging classifications indicated that 10.08% were at the T1 stage, 28.21% at the T2 stage, 25.08% at the T3 stage, and 36.63% at the T4 stage. Categorisation of nodal involvement showed that 23.38% were at stage N0, 9.71% were at stage N1, 46.80% were at stage N2, and 20.10% were at stage N3. For treatment, 56.88% of patients received chemotherapy, 75.93% underwent radiotherapy, while nearly all patients (94.34%) did not undergo primary lung tumor surgery. The metastasis prevalence was represented as follows: bone metastases have the highest rate of other distant organ metastases, accounting for 31.66% of all patients, followed by lung metastases (25.31%) and liver metastases (16.25%).

Univariable and multivariable Cox regression analysis

Univariable and multivariable Cox regression analyses were conducted to identify the independent factors related with both OS and LCSS of patients with LCBM (**Table 2**). We found that higher age, male gender, grade III/IV, stage T2-T4, stage N2-N3, other distant organ metastases were all significantly correlated with worse outcomes for both OS and LCSS. Months from diagnosis to therapy, chemotherapy, radiotherapy, primary lung tumor surgery were all favorable independent factors for patients with LCBM. C Compared with whites, survival was not different for blacks, but survival was better for other races, such as American Indian or Alaska Native, Asian, or Pacific Islander. In comparison to marital status, being divorced or separated, single and widowed were all associated with poorer OS, while being single did not affect LCSS. For pathological type, compared with adenocarcinoma, squamous cell, small cell, large cell and other carcinoma were all unfavorable independent factors for both OS and LCSS.

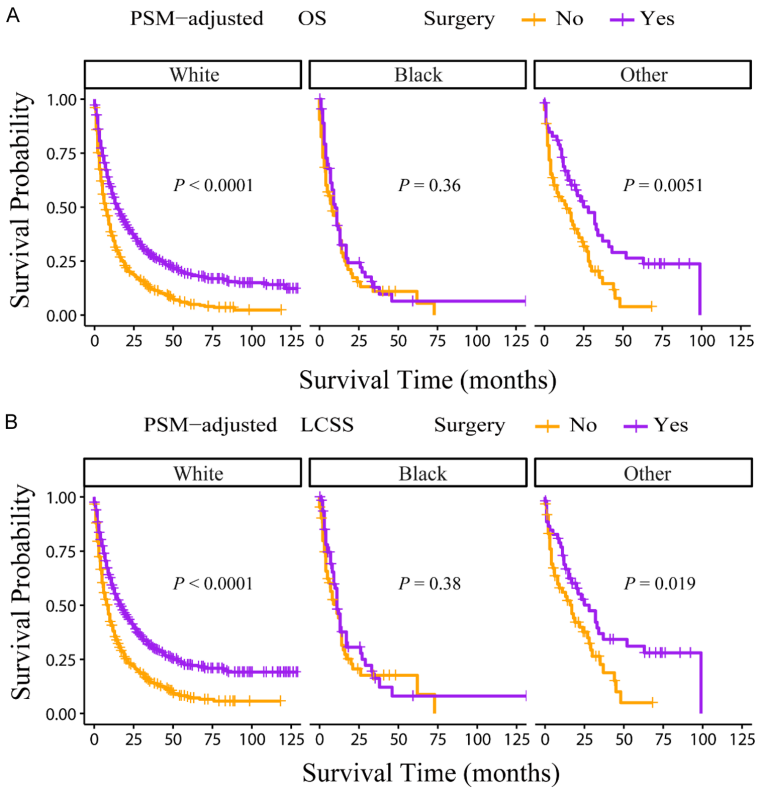


Figure 9. OS and LCSS of LCBM patients underwent primary lung tumor surgery or not after PSM adjustment (Stratified by race). A: OS of LCBM patients with different races; B: LCSS of LCBM patients with different races. PSM: propensity score matching, OS: overall survival, LCSS: lung cancer specific survival, LCBM: lung cancer brain metastases.

Constructing and assessing prognostic models of patients with LCBM

Given the above, patients were randomly assigned to training and test groups in a 7:3 ratio (**Table 1**) and the univariate and multivariate Cox analyses were performed again in the training data (Supplementary Table 2) to screen model features, and finally fifteen independent prognostic factors were selected as features of the models. We used seven machine learning algorithms to create prognostic models for assessing the OS of patients with LCBM at 6 months, 1 and 2-year. And then ROC curves of both training and test sets were plotted and their AUCs were calculated to compare the discrimination of these models.

Our XGBoost algorithm models manifested high discrimination in the survival prediction of patients with LCBM at 6-month (training data: AUC=0.820; test data: AUC=0.806), at 1-year (train data: AUC=0.806; test set: AUC=0.785), at 2-year (training data: AUC=0.807; test data: AUC=0.782) (**Figure 2A-F**). Compared with other machine learning algorithms, LR (6-month: AUC=0.777; 1-year: AUC=0.761; 2-year: AUC=0.765), ANN (6-month: AUC=0.775; 1-year: AUC=0.753; 2-year: AUC=0.751), RSF (6-month: AUC=0.757; 1-year: AUC=0.746; 2-year: AUC=0.726), DT (6-month: AUC=0.695; 1-year: AUC=0.716; 2-year: AUC=0.511), KNN (6-month: AUC=0.647; 1-year: AUC=0.611; 2-year: AUC=0.627), SVM (6-month: AUC=0.714; 1-year: AUC=0.593; 2-year: AUC=0.526), XGBoost algorithm models performed best (**Table 3**).

For further validation of our models, we collected clinical and prognostic information of 32 patients with LCBM at our hospital (Supplementary Table 3). In this external, independent data set, our XGBoost models still had a good level of robustness [6-month: AUC=0.848 (**Figure 3A**); 1-year: AUC=0.701 (**Figure 3B**); 2-year: AUC=0.810 (**Figure 3C**)].

Then, calibration curves of the training and test data were used to assess the accuracy of our XGBoost models (**Figure 4A-F**), the predicted values of XGBoost models were nearly in line with the observed values, indicating that XGBoost models had excellent accuracy. Meanwhile, decision curve analysis (DCA) was conducted to evaluate the clinical applicability of the models. The results showed that XGBoost models had a wide threshold probability range and well net benefit in predicting 6-month, 1-year and 2-year OS for LCBM (**Figure 5A-F**). Overall, the performance of our models was good.

In addition, we rated how prominent the clinical features were in the models. The results showed that the top 5 factors affecting prognosis were chemotherapy, pathological type, surgery, T stage and age (**Figure 6A-C**). Among these, chemotherapy was the most important factor in all three prognostic models (**Figure 6A-C**).

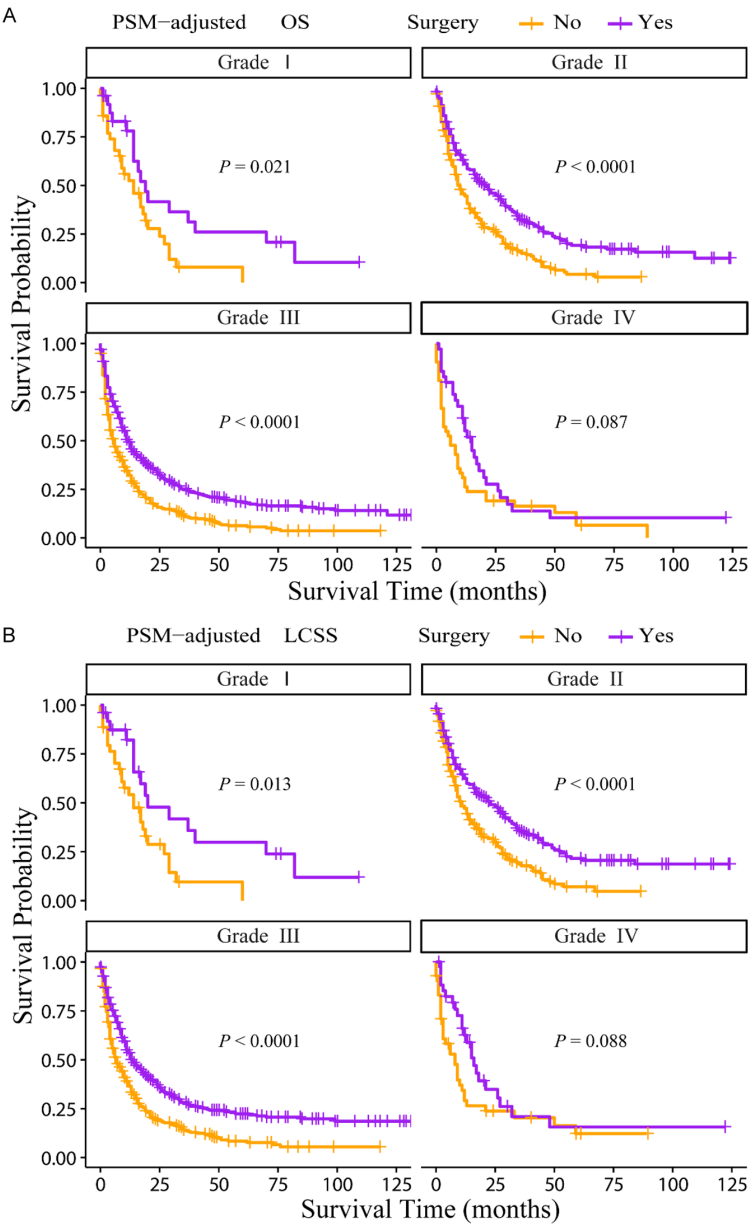


Figure 10. OS and LCSS of LCBM patients underwent primary lung tumor surgery or not after PSM adjustment (Stratified by grade). A: OS of LCBM patients with different grades; B: LCSS of LCBM patients with different grades. PSM: propensity score matching, OS: overall survival, LCSS: lung cancer specific survival, LCBM: lung cancer brain metastases.

Factors affecting the benefit of primary lung tumor surgery for patients with LCBM

Unexpectedly, primary lung tumor surgery was an independent prognostic factor for patients with LCBM in our multivariable Cox regression analysis (**Table 2**). Usually, this surgery is not recommended for LC patients with distant organ metastases, however, we found that 5.66% of patients with LCBM underwent the primary lung tumor surgery in this SEER cohort. Hence, we took a further look at which factors may affect the benefit of the surgery for patients with LCBM. A comparison of baseline features between patients underwent surgery and those without surgery revealed noticeable differences (**Table 4**). Therefore, PSM was used to help eliminate the disparity. After PSM adjustment, there were no differences in baseline characteristics between the two groups (**Table 4**).

According to the PSM-adjusted data, primary lung tumor surgery dramatically improved both OS (**Figure 7A**) and LCSS (**Figure 7B**) of patients with LCBM. Then, further stratification analyses were conducted to investigate the factors which may affect the benefit of the surgery. The patients < 70 years old could benefit from surgery for OS, while the patients < 80 years old could benefit for LCSS (**Figure 8A** and **8B**). Of all races, blacks could not benefit from surgery (**Figure 9A** and **9B**). In terms of pathological characteristics, there was no difference in prognosis between the surgery and no surgery groups for patients with grade IV (**Figure 10A** and **10B**), stage T4 (**Figure 11A** and **11B**), stage N3 (**Figure 12A** and **12B**), other distant organ metastases (**Figure 13A** and **13B**), squamous cell carcinoma or large cell carcinoma (**Figure 14A** and **14B**). In addition, it's very interesting that only patients who have received radiotherapy can benefit from the surgery (**Figure 15A** and **15B**). Other influencing factors were also analyzed, but they could not change the benefit of the surgery (data not show).

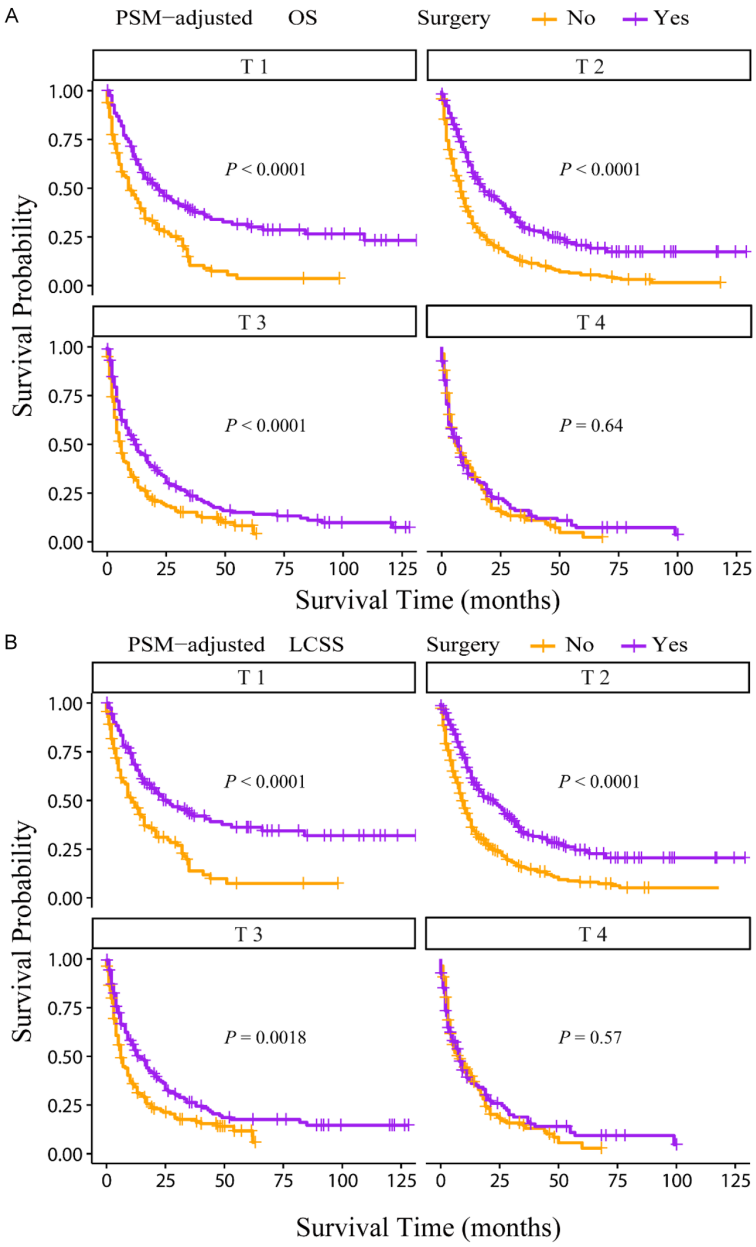


Figure 11. OS and LCSS of LCBM patients underwent primary lung tumor surgery or not after PSM adjustment (Stratified by T stage). A: OS of LCBM patients with different T stages; B: LCSS of LCBM patients with different T stages. PSM: propensity score matching, OS: overall survival, LCSS: lung cancer specific survival, LCBM: lung cancer brain metastases.

Discussion

The common metastasis sites of LC include the central nervous system, bone, liver, lung and adrenal gland [24, 25]. BM is not only the most common site of metastasis, but also the disease with the greatest impact on patients' prognosis and quality of life [26]. Although there has been a lot of progress in the treatment of lung cancer, the curative effect of BM is still limited. Thus, for these patients with LCBM and their families, survival time is their prime concern, we need prognostic prediction models to help them solve this issue. Unfortunately, accurate predictive models are lacking in the clinic. Using the SEER database, previous studies have built several nomograms to predict the prognosis of patients with LCBM, however, the accuracy rates of these models are all less than 70% [14-17] and these models can only work for some special conditions. There is no model which can be applied widely and accurately to predict OS of LCBM. To address this gap, we created several models by machine learning algorithms to predict 6-month, 1- and 2-year OS of patients with LCBM. To the best of our knowledge, the current study is the most up-to-date and largest dataset analysing the clinical characteristics and prognosis of patients with LCBM, and it is also the first one to create AI prognostic models with the highest accuracy for patients with LCBM. In addition, our XGBoost models performed well in an externally independent dataset, demonstrating their high clinical utility.

For the XGBoost machine learning algorithm, there are some fundamentals and benefits. Firstly, a decision tree is a tree structure similar to a flowchart, where the branch nodes represent a test on a feature, a classification based on the results of the test, and the leaf nodes represent a category. The principle of XGBoost is to add trees to a decision tree to reduce the difference (i.e. the loss function) between the predicted and actual results. For

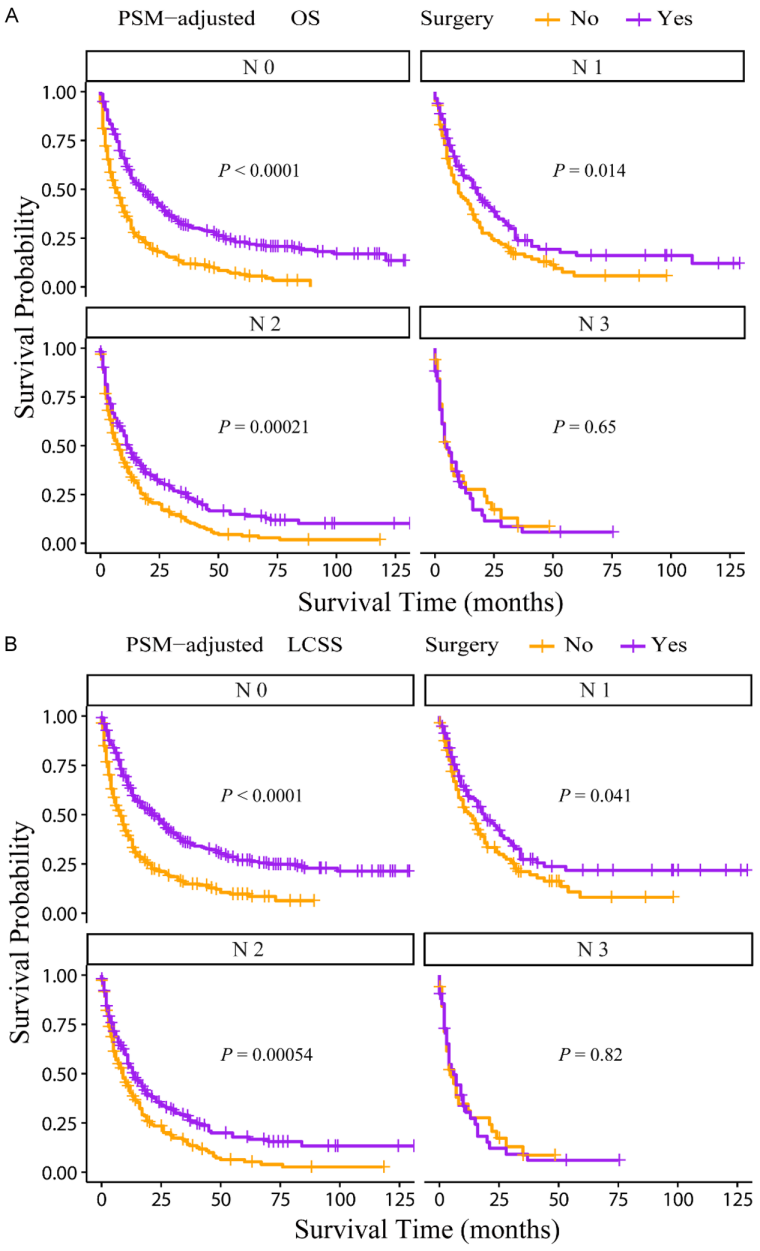


Figure 12. OS and LCSS of LCBM patients underwent primary lung tumor surgery or not after PSM adjustment (Stratified by N stage). A: OS of LCBM patients with different N stages; B: LCSS of LCBM patients with different N stages. PSM: propensity score matching, OS: overall survival, LCSS: lung cancer specific survival, LCBM: lung cancer brain metastases.

example, given a dataset with a number of cases, each case has m-dimensional features. By training the dataset, we obtain n trees. The cumulative value of these n trees is our predicted value. The accuracy of the algorithm increases as we add a tree to the n-1 trees to become n trees. But at the same time, if a tree reduces the loss function to a very low level, then there is less room for subsequent optimisation and eventually it will be prone to overfitting. Feature subsampling can be understood as selecting some of the m features to train each tree (similar to random forest) to improve the generalisation ability of the model and make it more diverse. In addition, compared to nomograms, XGBoost models can allow for missing parameters, but nomograms will not work if parameters are missing.

Several favorable independent prognostic factors of patients with LCBM were successfully identified in our study, including age < 50, female, non-whites and non-blacks, married or single, ≥ 1 month from diagnosis to therapy, adenocarcinoma, grade I, T1, N0 or N1, chemotherapy, radiotherapy, surgery and no other distant organ metastases. A recent study showed that patients with LCBM age < 40 years old tended to have a better OS [16], while other studies indicated the age > 66/70 years old was a risk factor in SCLC [14, 15], the age groups from the different studies were not the same, but the age groups in our groups were the most, and the trend was similar that older age had a worse prognosis.

In this cohort, the white people, black people and other people account for 77.20%, 12.15% and 10.65%, respectively, and most of other people are Asian or Pacific Islander. Compared with the white people, other people is a favorable independent prognostic factor for patients with LCBM, but we don't know how many people are Asian in the other people exactly, due to the demographic composition of the SEER database. Considering potential ethnic and genetic varia-

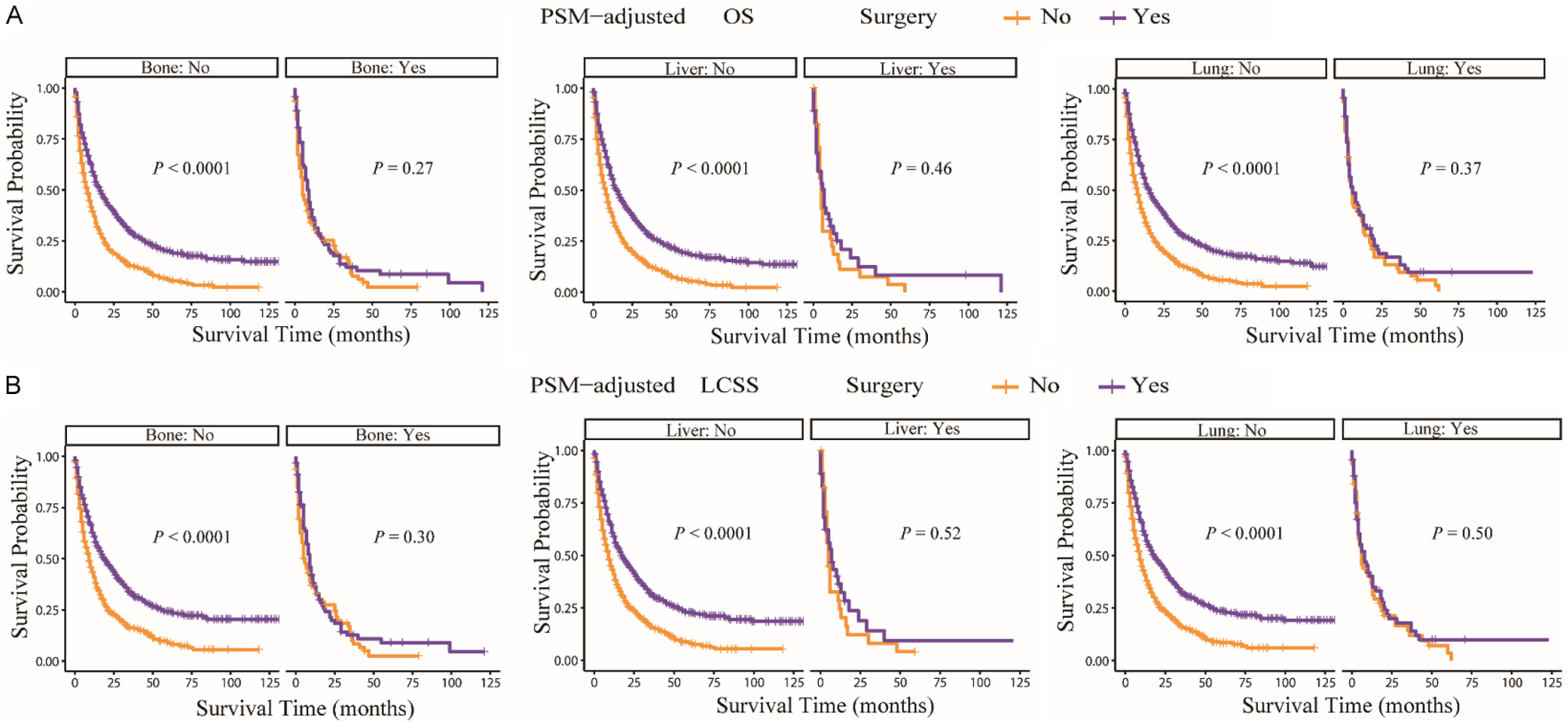


Figure 13. OS and LCSS of LCBM patients underwent primary lung tumor surgery or not after PSM adjustment (Stratified by distant organ metastases). A: OS of LCBM patients with different other distant organ metastases; B: LCSS of LCBM patients with different other distant organ metastases. PSM: propensity score matching, OS: overall survival, LCSS: lung cancer specific survival, LCBM: lung cancer brain metastases.

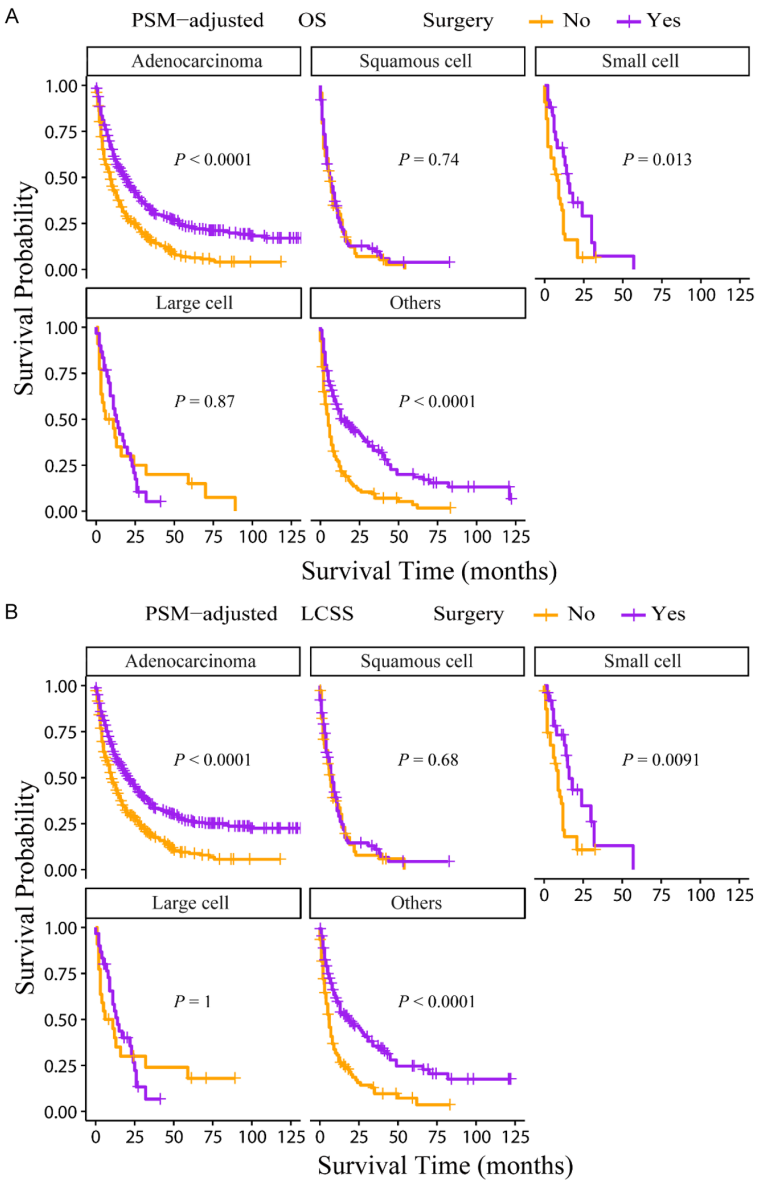


Figure 14. OS and LCSS of LCBM patients underwent primary lung tumor surgery or not after PSM adjustment (Stratified by pathological type). A: OS of LCBM patients with different pathological types; B: LCSS of LCBM patients with different pathological types. PSM: propensity score matching, OS: overall survival, LCSS: lung cancer specific survival, LCBM: lung cancer brain metastases.

tions in lung cancer prognosis, the applicability of the findings to non-American populations, especially Asian populations is uncertain. Most studies found that married patients had a better prognosis than unmarried patients [14, 15, 17], however, our study showed that the OS and LCSS of married patients were similar to those of single patients, and were better than those of divorced, separated or widowed patients. This is because these studies did not further stratify unmarried patients. Married cancer patients can receive practical support from their spouse (e.g. help with transport, paperwork, housework) to enable them to concentrate fully on their treatment. Importantly, a partner can provide emotional support, which can reduce the stress of cancer treatment [27]. The effect of partner social support may be physiologically mediated through neuroendocrine, nervous and immune interactions directly related to cancer [28]. The divorced, separated or widowed patients may desire for husband's help, but the broken marital relationships might have exacerbated the patients' psychological trauma, leading to poorer prognosis. While single patients did not have these mental injuries and they also could get help from their families, so the differences in survival between married and single patients are small. In this study, we investigated for the first time the role of time from diagnosis to treatment on prognosis of patients with LCBM and found that ≥ 1 month from diagnosis to therapy was a favorable independent prognostic factor. This does not mean that the later treatment is better, but it may be that waiting for more comprehensive examination and possible clinical trials will benefit for these patients.

For the treatment, we know that systematic treatment can improve the prognosis of patients with LCBM [25, 26, 29, 30], however, to our surprise, we found that in addition to chemotherapy and radiotherapy, primary lung tumor surgery was also an independent prognostic factor, which is usually not recommended for stage IV patients [25, 31, 32]. Some studies indicated that for some oligometastatic LC (resectable N0, 1 primary), the surgery for

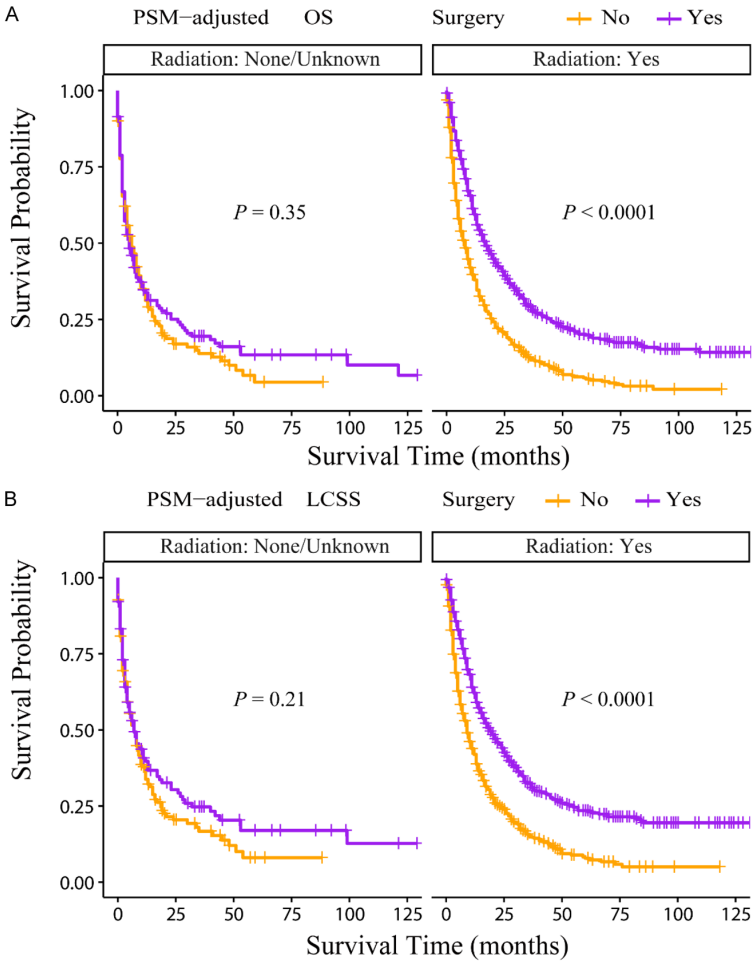


Figure 15. OS and LCSS of LCBM patients underwent primary lung tumor surgery or not after PSM adjustment (Stratified by radiotherapy). A: OS of LCBM patients with radiotherapy or not; B: LCSS of LCBM patients with radiotherapy or not radiotherapy. PSM: propensity score matching, OS: overall survival, LCSS: lung cancer specific survival, LCBM: lung cancer brain metastases.

primary lung tumor and surgery or radiosurgical treatment for synchronous brain metastases might prolonged the prognosis [33-36]. This may explain only 5.66% patients with LCBM received primary lung tumor in our cohort, but there is no data about the number of brain metastases in the SEER database and there were still some patients with multiple organ metastases underwent primary lung surgery. Thus, we further conducted stratified analyses to investigate the factors affecting therapeutic effect of the primary lung tumor surgery in patients with LCBM, which were never reported. We found that age > 70 years old, blacks, grade IV, stage T4, N3, other distant organ metastases, squamous cell carcinoma, large cell carcinoma and no radiation were all unfavorable factors of primary lung tumor surgery for the prognosis of patients with LCBM. Of course, we need further clinical trials to confirm these results, but our study provides a lot of data reference for future treatment management and clinical trials.

Despite its promising findings, our study has some limitations. Firstly, SEER data may be a good representation of the general situation, but due to ethnic differences, it may not always be the case for Asians, particularly Chinese. Secondly, this is a retrospective study, there may be selection bias during the screening of patients and PSM. Thirdly, some detailed treatment information is not available from the SEER database, such as targeted drugs, chemotherapeutic drugs, immunotherapy and radiotherapy site, especially EGFR/ALK mutation information, which are also important variables for patients with LCBM, these unavailable variables may introduce biases. Fourthly, all these patients are de novo LCBM, we can not get the data of patients who later developed brain metastases after being diagnosed with LC from SEER database. Fifthly, our study includes external validation using a small sample from a single hospital, the robustness of the models would benefit from further validation using larger, more diverse datasets, potentially from multiple institutions or regions.

In the future, the models that constantly integrate real-time data and continue to adjust and optimize themselves will enhance their accuracy and applicability. Meanwhile, prospective studies about the surgery on primary lesion for the prognosis of LCBM need to confirm the findings.

Conclusions

We extensively investigated the clinical characteristics and prognosis of patients with LCBM and created some machine learning models to predict their 6 month, 1-year and 2-year OS. Compared to traditional nomograms, these machine learning models offer more accurate predictions for patients and doctors. Furthermore, our in-depth stratified analysis showed that age > 70 years old, blacks, grade IV, stage T4, N3, other distant organ metastases, squamous cell carcinoma, large cell carcinoma and no radiation were all unfavorable factors of primary lung tumor surgery for the prognosis of patients with LCBM.

Acknowledgements

This work was funded in part by the following: National Science Foundation of China (8237-4229 to X. Zhao; 82103569 to J.K. Qu); Free exploring fund of Xi'an Jiaotong University (xzy012022096 to X. Zhao; xzy012022097 to J.K. Qu). Medical "basic - clinical" integration and innovation project of Xi'an Jiaotong University (YXJLRH2022088 to J.K. Qu). Key laboratory construction project of tumor prevention and treatment of integrated Chinese and western medicine in Shaanxi province (2022-ZXY-SYS-002 to S.Q. Zhang); and Young Technology Star Project (2024ZC-KJXX-083 to X. Zhao).

Disclosure of conflict of interest

None.

Address correspondence to: Shuqun Zhang and Jingkun Qu, Department of Surgical Oncology, The Comprehensive Breast Care Center, The Second Affiliated Hospital of Xi'an Jiaotong University, 157 West Fifth Road, Xi'an 710004, Shaanxi, P. R. China. E-mail: shuqun_zhang1971@163.com (SQZ); qujingkun@xjtu.edu.cn (JKQ)

References

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

[2] Siegel RL, Miller KD, Wagle NS and Jemal A. Cancer statistics, 2023. CA Cancer J Clin 2023; 73: 17-48.

[3] Qi J, Li M, Wang L, Hu Y, Liu W, Long Z, Zhou Z, Yin P and Zhou M. National and subnational trends in cancer burden in China, 2005-20: an analysis of national mortality surveillance data. Lancet Public Health 2023; 8: e943-e955.

[4] Duma N, Santana-Davila R and Molina JR. Non-small cell lung cancer: epidemiology, screening, diagnosis, and treatment. Mayo Clin Proc 2019; 94: 1623-1640.

[5] Zhang X, Gao H, Dang S, Dai L and Zhang J. Extracranial metastasis sites correlate to the incidence risk of brain metastasis in stage IV non-small cell lung cancer: a population-based study. J Cancer Res Clin Oncol 2023; 149: 6293-6301.

[6] Rudin CM, Brambilla E, Faivre-Finn C and Sage J. Small-cell lung cancer. Nat Rev Dis Primers 2021; 7: 3.

[7] Pan K, Concannon K, Li J, Zhang J, Heymach JV and Le X. Emerging therapeutics and evolving assessment criteria for intracranial metastases in patients with oncogene-driven non-small-cell lung cancer. Nat Rev Clin Oncol 2023; 20: 716-732.

[8] Zhang Q, Abdo R, Iosef C, Kaneko T, Cecchini M, Han VK and Li SS. The spatial transcriptomic landscape of non-small cell lung cancer brain metastasis. Nat Commun 2022; 13: 5983.

[9] Hao Y and Li G. Risk and prognostic factors of brain metastasis in lung cancer patients: a surveillance, epidemiology, and end results population-based cohort study. Eur J Cancer Prev 2023; 32: 498-511.

[10] Ulahannan D, Khalifa J, Faivre-Finn C and Lee SM. Emerging treatment paradigms for brain metastasis in non-small-cell lung cancer: an overview of the current landscape and challenges ahead. Ann Oncol 2017; 28: 2923-2931.

[11] Karimpour M, Ravanbakhsh R, Maydanchi M, Rajabi A, Azizi F and Saber A. Cancer driver gene and non-coding RNA alterations as biomarkers of brain metastasis in lung cancer: a review of the literature. Biomed Pharmacother 2021; 143: 112190.

[12] Liu W, Powell CA and Wang Q. Tumor microenvironment in lung cancer-derived brain metastasis. Chin Med J (Engl) 2022; 135: 1781-1791.

[13] Peters S, Bexelius C, Munk V and Leighl N. The impact of brain metastasis on quality of life, resource utilization and survival in patients with non-small-cell lung cancer. Cancer Treat Rev 2016; 45: 139-162.

[14] Rong YT, Zhu YC and Wu Y. A novel nomogram predicting cancer-specific survival in small cell

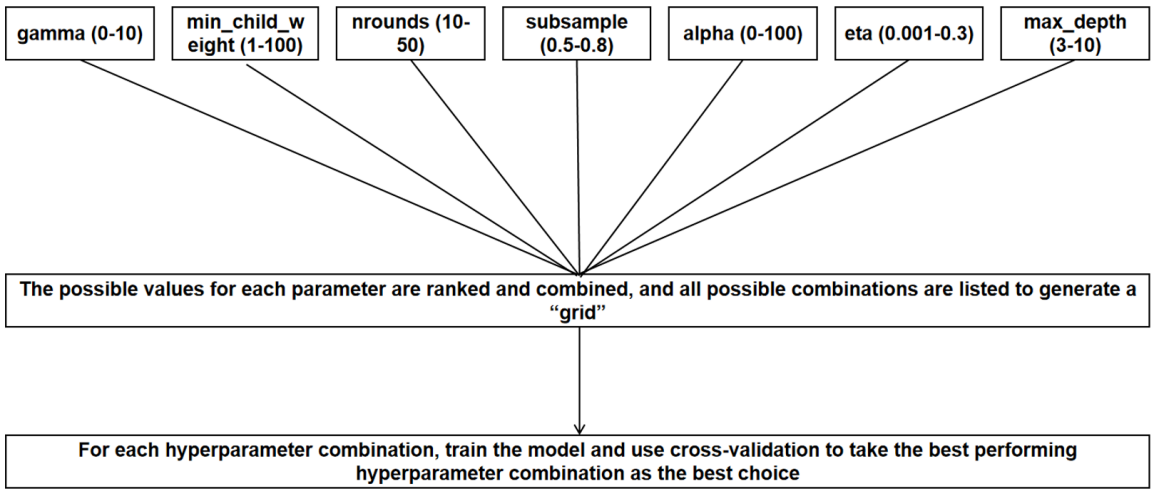
lung cancer patients with brain metastasis. Transl Cancer Res 2022; 11: 4289-4302.
[15] Shan Q, Shi J, Wang X, Guo J, Han X, Wang Z and Wang H. A new nomogram and risk classification system for predicting survival in small cell lung cancer patients diagnosed with brain metastasis: a large population-based study. BMC Cancer 2021; 21: 640.
[16] Zhang GH, Liu YJ and De Ji M. Risk factors, prognosis, and a new nomogram for predicting cancer-specific survival among lung cancer patients with brain metastasis: a retrospective study based on SEER. Lung 2022; 200: 83-93.
[17] Yuan J, Cheng Z, Feng J, Xu C, Wang Y, Zou Z, Li Q, Guo S, Jin L, Jiang G, Shang Y and Wu J. Prognosis of lung cancer with simple brain metastasis patients and establishment of survival prediction models: a study based on real events. BMC Pulm Med 2022; 22: 162.
[18] Liang M, Chen M, Singh S and Singh S. Construction, validation, and visualization of a web-based nomogram to predict overall survival in small-cell lung cancer patients with brain metastasis. Cancer Causes Control 2024; 35: 465-475.
[19] Li C, Liu M, Li J, Wang W, Feng C, Cai Y, Wu F, Zhao X, Du C, Zhang Y, Wang Y, Zhang S and Qu J. Machine learning predicts the prognosis of breast cancer patients with initial bone metastases. Front Public Health 2022; 10: 1003976.
[20] Li C, Liu M, Zhang Y, Wang Y, Li J, Sun S, Liu X, Wu H, Feng C, Yao P, Jia Y, Zhang Y, Wei X, Wu F, Du C, Zhao X, Zhang S and Qu J. Novel models by machine learning to predict prognosis of breast cancer brain metastases. J Transl Med 2023; 21: 404.
[21] Qu J, Li C, Liu M, Wang Y, Feng Z, Li J, Wang W, Wu F, Zhang S and Zhao X. Prognostic models using machine learning algorithms and treatment outcomes of occult breast cancer patients. J Clin Med 2023; 12: 3097.
[22] Kane LT, Fang T, Galetta MS, Goyal DKC, Nicholson KJ, Kepler CK, Vaccaro AR and Schroeder GD. Propensity score matching: a statistical method. Clin Spine Surg 2020; 33: 120-122.
[23] Barakat A, Mittal A, Ricketts D and Rogers BA. Understanding survival analysis: actuarial life tables and the Kaplan-Meier plot. Br J Hosp Med (Lond) 2019; 80: 642-646.
[24] Riihimaki M, Hemminki A, Fallah M, Thomsen H, Sundquist K, Sundquist J and Hemminki K. Metastatic sites and survival in lung cancer. Lung Cancer 2014; 86: 78-84.
[25] Zhu Y, Cui Y, Zheng X, Zhao Y and Sun G. Small-cell lung cancer brain metastasis: from molecular mechanisms to diagnosis and treatment. Biochim Biophys Acta Mol Basis Dis 2022; 1868: 166557.
[26] Wang B, Guo H, Xu H, Yu H, Chen Y and Zhao G. Research progress and challenges in the treatment of central nervous system metastasis of non-small cell lung cancer. Cells 2021; 10: 2620.
[27] de Graeff A, de Leeuw JR, Ros WJ, Hordijk GJ, Blijham GH and Winnubst JA. Sociodemographic factors and quality of life as prognostic indicators in head and neck cancer. Eur J Cancer 2001; 37: 332-339.
[28] Lutgendorf SK, Sood AK, Anderson B, McGinn S, Maiseri H, Dao M, Sorosky JI, De Geest K, Ritchie J and Lubaroff DM. Social support, psychological distress, and natural killer cell activity in ovarian cancer. J Clin Oncol 2005; 23: 7105-7113.
[29] Singh SA, McDermott DM and Mattes MD. Impact of systemic therapy type and timing on intracranial tumor control in patients with brain metastasis from non-small-cell lung cancer treated with stereotactic radiosurgery. World Neurosurg 2020; 144: e813-e823.
[30] Frega S, Bonanno L, Guarneri V, Conte P and Pasello G. Therapeutic perspectives for brain metastases in non-oncogene addicted non-small cell lung cancer (NSCLC): towards a less dismal future? Crit Rev Oncol Hematol 2018; 128: 19-29.
[31] Singh N, Jaiyesimi IA, Ismaila N, Leighl NB, Mamdani H, Phillips T and Owen DH. Therapy for stage IV non-small-cell lung cancer with driver alterations: ASCO living guideline, version 2023.1. J Clin Oncol 2023; 41: e42-e50.
[32] Singh N, Jaiyesimi IA, Ismaila N, Leighl NB, Mamdani H, Phillips T and Owen DH. Therapy for stage IV non-small-cell lung cancer without driver alterations: ASCO living guideline, version 2023.1. J Clin Oncol 2023; 41: e51-e62.
[33] Endo C, Hasumi T, Matsumura Y, Sato N, Deguchi H, Oizumi H, Sagawa M, Tsushima T, Takahashi S, Shibuya J, Hirose M and Kondo T. A prospective study of surgical procedures for patients with oligometastatic non-small cell lung cancer. Ann Thorac Surg 2014; 98: 258-264.
[34] Antuna AR, Vega MA, Sanchez CR and Fernandez VM. Brain metastases of non-small cell lung cancer: prognostic factors in patients with surgical resection. J Neurol Surg A Cent Eur Neurosurg 2018; 79: 101-107.
[35] Kozower BD, Larner JM, Detterbeck FC and Jones DR. Special treatment issues in non-small cell lung cancer: diagnosis and management of lung cancer, 3rd ed: american college

of chest physicians evidence-based clinical practice guidelines. Chest 2013; 143 Suppl: e369S-e399S.

[36] Gray PJ, Mak RH, Yeap BY, Cryer SK, Pinnell NE, Christianson LW, Sher DJ, Arvold ND, Baldini EH, Chen AB, Kozono DE, Swanson SJ, Jackman DM and Alexander BM. Aggressive therapy for patients with non-small cell lung carcinoma and synchronous brain-only oligometastatic disease is associated with long-term survival. Lung Cancer 2014; 85: 239-244.

Supplementary Table 1. Main parameters of the XGBoost model

Parameter	Value
Gamma	1
Scale_pos_weight	1
Min_child_weight	5
Subsample	0.7
Alpha	1
Nround	20
Max_depth	5
Eta	0.1



Supplementary Figure 1. The flowchart detailed the procedure for main hyperparameters selection of XGBoost models. We used "GridSearch" ridSearch s.GBoost models.XGBoost models.dhyperparameters (including gamma, alpha, max depth, nrounds, eta, min_child_weight, etc.). The possible values for each parameter are ranked and combined, and all possible combinations are listed to generate a "grid". For each hyperparameter combination, train the model and use cross-validation to take the best performing hyperparameter combination as the best choice.

Supplementary Table 2. Univariate and multivariate Cox analysis of LCBM characteristics extracted from training data

	OS					
	Univariate Cox analysis			Multivariate Cox analysis		
	HR	95% CI	*P* Value	HR	95% CI	*P* Value
Age						
< 50	Reference			Reference		
50-59	1.33	1.20-1.47	***	1.21	1.08-1.35	***
60-69	1.50	1.36-1.66	***	1.33	1.19-1.47	***
70-79	1.88	1.69-2.08	***	1.61	1.44-1.80	***
≥ 80	2.41	2.14-2.73	***	1.68	1.47-1.93	***
Sex						
Female	Reference			Reference		
Male	1.29	1.23-1.34	***	1.23	1.17-1.30	***
Race						
White	Reference			Reference		
Black	1.07	0.99-1.14	0.07	1.00	0.93-1.09	0.89
Others	0.68	0.63-0.74	***	0.71	0.65-0.77	***
Marriage status						
Married	Reference			Reference		
Divorced/Separated	1.28	1.20-1.37	***	1.24	1.15-1.33	***
Single	1.16	1.09-1.23	***	1.07	1.00-1.14	0.07
Widowed	1.33	1.24-1.43	***	1.12	1.03-1.22	**
Months from diagnosis to therapy						
0 month	Reference			Reference		
≥ 1 month	0.82	0.78-0.86	***	0.78	0.74-0.82	***
Pathological type						
Adenocarcinoma	Reference			Reference		
Squamous cell	1.71	1.60-1.83	***	1.48	1.37-1.59	***
Small cell	1.37	1.27-1.48	***	1.31	1.18-1.46	***
Large cell	1.29	1.11-1.51	**	1.38	1.17-1.64	***
Others	1.50	1.41-1.59	***	1.30	1.21-1.40	***
Grade						
Grade I	Reference			Reference		
Grade II	1.06	0.92-1.21	0.42	1.14	0.98-1.31	0.08
Grade III	1.42	1.24-1.61	***	1.36	1.18-1.56	***
Grade IV	1.52	1.31-1.76	***	1.39	1.17-1.65	***
T Stage						
T1	Reference			Reference		
T2	1.31	1.20-1.42	***	1.34	1.23-1.47	***
T3	1.53	1.41-1.67	***	1.50	1.37-1.65	***
T4	1.57	1.45-1.71	***	1.50	1.37-1.65	***
N Stage						
N0	Reference			Reference		
N1	1.06	0.97-1.16	0.18	1.05	0.95-1.15	0.36
N2	1.19	1.12-1.26	***	1.15	1.08-1.23	***
N3	1.23	1.14-1.31	***	1.18	1.09-1.28	***
Chemotherapy						
No/unknown	Reference			Reference		
Yes	0.36	0.34-0.38	***	0.39	0.37-0.41	***

Radiotherapy						
No/unknown	Reference			Reference		
Yes	0.59	0.56-0.62	***	0.92	0.86-0.99	*
Surgery						
No	Reference			Reference		
Yes	0.48	0.43-0.53	***	0.54	0.48-0.60	***
Bone metastasis						
No	Reference			Reference		
Yes	1.14	1.09-1.20	***	1.20	1.13-1.27	***
Liver metastasis						
No	Reference			Reference		
Yes	1.46	1.37-1.55	***	1.34	1.25-1.44	***
Lung metastasis						
No	Reference			Reference		
Yes	1.24	1.17-1.3	***	1.13	1.06-1.20	***

*P < 0.05, **P < 0.01, ***P < 0.001.

Supplementary Table 3. Relevant characteristics of LCBM patients included from our hospital for XGBoost models external validation

Age	Sex	Marital	Race	Months	Pathological	Grade	T	N	Surgery	Radiation	Chemotherapy	Bone	Liver	Lung	6-month status	1-year status	2-year status
53	1	1	3	1	1	NA	NA	3	0	0	1	0	0	0	1	0	0
62	1	1	3	1	NA	NA	NA	3	0	0	1	1	1	1	0	0	0
56	2	1	3	1	3	NA	NA	3	1	0	1	0	0	0	1	1	0
69	2	1	3	1	1	NA	1	3	0	0	1	1	0	0	1	0	0
60	2	1	3	1	1	2	NA	NA	1	1	1	0	0	0	1	1	1
44	2	1	3	1	3	NA	NA	NA	0	1	0	0	0	0	0	0	0
63	2	1	3	1	2	NA	1	3	0	1	1	0	0	0	1	0	0
58	2	1	3	1	1	NA	NA	NA	0	1	1	0	0	0	1	0	0
61	2	1	3	1	NA	NA	NA	3	0	0	0	1	0	1	0	0	0
56	2	1	3	1	1	NA	NA	1	1	0	1	0	0	0	1	1	0
61	2	1	3	1	3	NA	NA	3	0	0	1	0	1	0	1	0	0
47	2	1	3	1	NA	NA	NA	NA	0	0	1	0	1	1	0	0	0
44	2	1	3	1	2	NA	4	3	0	0	0	0	0	1	0	0	0
45	1	1	3	1	3	NA	NA	3	0	1	1	0	0	0	1	0	0
72	2	1	3	1	NA	NA	3	NA	1	1	1	0	0	0	1	0	0
59	1	1	3	1	1	NA	4	NA	0	0	1	0	0	0	0	0	0
66	2	1	3	1	1	NA	NA	3	0	0	0	0	0	0	0	0	0
42	1	1	3	1	1	NA	1	3	0	0	1	0	0	0	0	0	0
75	1	4	3	1	1	3	NA	NA	0	1	0	0	0	0	0	0	0
62	2	1	3	1	2	NA	2	NA	0	1	1	1	0	1	0	0	0
54	2	1	3	1	5	1	NA	3	0	0	1	0	0	1	1	0	0
60	2	1	3	1	1	NA	4	0	0	0	1	1	1	1	0	0	0
55	1	1	3	1	1	NA	NA	1	1	1	1	1	0	0	1	1	1
69	2	1	3	1	5	NA	NA	NA	1	0	1	1	0	0	1	0	0
61	2	1	3	1	1	NA	NA	0	1	1	1	1	0	0	1	1	1
54	2	1	3	1	5	3	NA	0	1	0	0	0	0	0	0	0	0
72	2	1	3	1	2	3	NA	NA	0	0	0	0	0	0	0	0	0
56	2	1	3	2	2	NA	NA	NA	0	0	1	0	0	0	1	1	0
73	2	1	3	1	NA	NA	NA	NA	1	1	1	1	0	0	1	0	0
64	2	1	3	1	3	NA	NA	NA	1	0	1	0	0	0	1	0	0
52	2	1	3	1	3	NA	NA	NA	0	0	1	0	0	0	1	1	0
49	2	1	3	1	3	NA	NA	NA	0	0	1	0	0	0	1	1	0