The construction and validation of the model for predicting the incidence and prognosis of brain metastasis in lung cancer patients

Chunjian Zuo
The First Affiliated Hospital of Chongqing Medical University

Peng Sun
The Second Affiliated Hospital of Harbin Medical University National Cancer Center/Chinese Academy of Medical Sciences and Peking Union Medical College

Guanchu Liu
The First Affiliated Hospital of Chongqing Medical University

Pengyu Che
The people's Hospital of Chongqing Hechuan

Gang Li
Chonggang General Hospital

Xiang Li
The people's Hospital of Chongqing Tongnan

Ye Bai
School of Public Health and Management, Chongqing Medical University

Jie Tian
The Third Affiliated Hospital of Chongqing Medical University (Gener Hospital)

Huanwen Chen (coolstarchw9527@163.com)
The First Affiliated Hospital of Chongqing Medical University

Research

Keywords: Lung cancer, Brain metastases, Incidence, Prognosis, Prediction model

Posted Date: February 6th, 2020

DOI: https://doi.org/10.21203/rs.2.22779/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Version of Record: A version of this preprint was published at Translational Cancer Research on January 1st, 2021. See the published version at https://doi.org/10.21037/tcr-20-2745.
Abstract

Background: Brain metastasis (BM) causes high morbidity and mortality rate in lung cancer (LC). The present study aims to develop models for predicting the development and prognosis of brain metastasis using a large sample size lung cancer cohort. Methods: A total of 266,522 lung cancer cases that were diagnosed between 2010 and 2016 were selected from the Surveillance, Epidemiology, and End Results Program (SEER) cohort. The risk factors for developing BM and prognosis were calculated by uni and multivariable logistic and Cox regression analysis, respectively and nomograms were constructed basing on the risk factors. The performance of the nomogram was evaluated by receiver operating characteristics curve (ROC) or C-index and calibration curve, respectively. Results: The prevalence of BM was 16.25%, the associated factors for developing BM including advanced age, Asian or Pacific Islander race, uninsured status, primary tumor site, higher T stage, N stage, poorly differentiated grade, the presence of lung, liver and bone metastases and adenocarcinoma histology. The median overall survival (OS) was 4 months; the associated prognosis factors were familiar with risk factors plus female gender, unmarried status, and surgery. The calibration curve showed good agreement between predicted and actual probability and the AUC/C-index were 73.1% (95% CI: 72.6-73.6%) and 0.88 (95 % CI 0.87-0.89) for risk and prognosis predictive models, respectively.

Background

As indicated in the latest in the GLOBOCAN report, about 2.1 million new diagnosed lung cancer (LC) cases and 1.8 million deaths were predicted worldwide in 2018, which ranks the leading cause of cancer morbidity and mortality [1]. Brain metastasis (BM) was the most prevalent intracranial tumors in adults, which commonly developed in LC patients [3-5]. BM generally shows poor prognosis, approximately 90% of patients died within two years after the initial diagnosis and the median survival is ranged from 7–10 months that significantly worsen the patients’ health and quantity of life [6, 7]. [2, 3]

It is urgent to identify the high-risk patients with BM and conduct targeted screening; however, National Comprehensive Cancer Network (NCCN) suggest that brain MRI should just be recommended for patients diagnosed with stage II to IV and high-risk stage 1B non-small-cell lung cancer and no unanimous screening guideline for identifying the BM exists until now [8, 9]. Previous studies have found several risk and prognostic factors for BM, which shed a light on the method to identify the high-risk patients, predict the survival of the patients and conduct targeted therapy [10-13]. However, due to the relatively lower sample size, the results were not consistent [7, 14, 15].

Surveillance, Epidemiology, and End Results (SEER) is an open-access database comprising approximately 30% of the total US population, which collected complete the social demographics and clinical data up to date for assessment of characteristics of cancer epidemiology worldwide.

The present study aims to investigate the prevalence of BM in the LC patient and the risk and prognostic factors for which, then to construct the nomograms for predicting the development and prognosis of the BM in LC patients using the SEER database.
Method

Ethics statement

Cancer information in the SEER database is continually reported in every state of the United States and retrieved with no need for informed patient consent. Our study complied with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Data source

Data in this population-based study were abstracted from the SEER. SEER*Stat Software version 8.3.4 (https://seer.cancer.gov/seerstat/) (Information Management Service, Inc. Calverton, MD, USA) was used to generate the case listing.

Cohort selection

A total of 266,522 eligible LC patients were identified between January 1, 2010, and December 31, 2016. The inclusion criteria were as follows: (1) age 18 and above; (2) diagnosed as the first and only cancer; (3) single primary site; (4) not diagnosed from a death certificate or an autopsy; (5) with definite BM information at primal diagnosis of LC. The flow-chart for the cohort selection was shown in Supplementary Fig. 1.

Statistical analysis

Absolute numbers and incidence proportions were calculated for patients with LC and BM identified at diagnosis with the classification strategies as: age (<41, 41–50, 51–60, 61–70, 71–80, 81–90 and >90 years), sex (female and male), race [white, black, AI (American Indian/Alaska Native), and API (Asian or Pacific Islander)], marital status (married and unmarried), insurance status (insured and uninsured), primary tumor site (Main bronchus, Upper lobe, Middle lobe, Lower lobe, Overlapping lesion of lung, Lung, NOS), primary tumor T stage (T1, T2, T3, and T4), regional lymph node stage (N0, N1, N2 and N3), tumor metastases stage (M0 and M1), tumor differentiated grade (grade I, grade II, grade III, and grade IV), the presence or absence of lung, liver, or bone metastases, histology [AC (Adenocarcinoma), SLC (Squamous lung cancer), ASC (Adenosquamous carcinoma), LCLC (Large cell lung cancer), SCLC (Small cell lung cancer) and Others]. All the data of “Unknown” in every categorical variable were included. The univariable and multivariable logistic regression model was conducted to determine the associated factors for developing BM.

The Kaplan–Meier method was used to analyze the overall survival (defined as the time from diagnosis of LC with BM to all causes of death) of the LC patients with BM. Meanwhile, the surgical treatment of primary site (not or yes) was further added with significant risk factors to identify the prognostic factors for LC patients with BM by using multivariable Cox proportional hazards regression.

Based on the results of multivariable logistic analysis and multivariable Cox proportional hazards regression, predictive nomograms were formulated using the rms package in R version 3.6.0 (R Foundation for Statistical Computing, Vienna, Austria; www.r-project.org). The receiver operating characteristics (ROC) and concordance index (C-index) were utilized to evaluate the discrimination of nomograms severally.
The calibration of the nomograms was assessed by calibration plots, which was generated by adopting a regression smoothing method and bootstrapping with 100 resamples to estimate the calibration of nomograms, where the consistency between observed and predicted probabilities of occurrence of BM and 1,3 and 5-year survival was presented graphically.

The randomly splitting method was adopted as the method of internal validation to evaluate the stability of the nomograms. The ROC curve was created to evaluate the performance of the construction set and validation set respectively, and the difference of the AUC was tested by DeLong’s test. The external validation set that consisted of LC patients with de novo BM in 2016 was used to evaluate the external performance of survival predicted nomogram. Statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS) version 23.0 software package for Windows (SPSS, Inc.). Statistically significant levels were set two-tailed at $P < 0.05$.

**Results**

**Incidence of BM**

Totally, 205,141 eligible patients were incorporated into the present study and 33,332 (16.25% of the entire cohort) subjects were determined with BM and 28,526 LC patients with BM were followed up for more than 1 year (Table 1). The prevalence of BM was increased from 18 years, peaked at 50 years and then was gradually decreased (Fig. 1A). The variation for the prevalence of BM with age in metastatic LC patients showed similar trends.
Table 1
Baseline of the demographic and related clinical characteristics for patients diagnosed with initial lung carcinoma.

| Subject characteristics | Patients’ No. of LC(2010–2016) | Patients’ No. of LC(2010–2015) |
|-------------------------|---------------------------------|---------------------------------|
|                         | With BM (N, %)                  | Without BM (N, %)               | With BM (N, %)                  | Without BM (N, %)               |
| Age                     |                                 |                                 |                                 |
| < 41                    | 363 (17.0)                      | 1775(83.0)                      | 303(16.5)                      | 1530(83.5)                      |
| 41–50                   | 2285(21.2)                      | 8494(78.8)                      | 2011(21.1)                     | 7541(78.9)                      |
| 51–60                   | 8841(19.1)                      | 37429(80.9)                     | 7586(19.1)                     | 32202(80.9)                     |
| 61–70                   | 11567(14.5)                     | 68374(85.5)                     | 9848(14.4)                     | 58371(85.6)                     |
| 71–80                   | 7601(10.4)                      | 65576(89.6)                     | 6479(10.4)                     | 55723(89.6)                     |
| 81–90                   | 2491(7.3)                       | 31794(92.7)                     | 2145(7.3)                      | 27366(92.7)                     |
| > 90                    | 184(5.2)                        | 3367(94.8)                      | 154(5.2)                       | 2821(94.8)                      |
| Sex                     |                                 |                                 |                                 |
| Male                    | 17411(13.4)                     | 112307(86.6)                    | 14907(13.4)                    | 96443(86.6)                     |
| Female                  | 15921(13.2)                     | 104502(86.8)                    | 13619(13.3)                    | 89111(86.7)                     |
| Race                    |                                 |                                 |                                 |
| White                   | 26015(12.9)                     | 175568(87.1)                    | 22400(13.0)                    | 150446(87.0)                    |
| Black                   | 4290(14.5)                      | 25223(85.5)                     | 3619(14.3)                     | 21647(85.7)                     |
| AI                      | 193(13.8)                       | 1201(86.2)                      | 159(13.5)                      | 1018(86.5)                      |
| API                     | 2764(16.3)                      | 14235(83.7)                     | 2296(16.0)                     | 12030(84.0)                     |
| Unknown                 | 70(10.7)                        | 582(89.3)                       | 52(11.2)                       | 413(88.8)                       |
| Marital status          |                                 |                                 |                                 |
| Others                  | 15401 (13.2)                    | 101707(86.8)                    | 13123(13.1)                    | 86870(86.9)                     |
| Married                 | 16557 (13.6)                    | 104930(86.4)                    | 14238(13.7)                    | 89838(86.3)                     |
| Unknown                 | 1374 (11.9)                     | 10172(88.1)                     | 1165(11.6)                     | 8846(88.4)                      |
| Insurance status        |                                 |                                 |                                 |

Abbreviations: LC = lung carcinoma; BM = brain metastases; Met = metastases; AI = American Indian/Alaska Native; API = Asian or Pacific Islander; MB = Main bronchus; UL = Upper lobe; ML = Middle lobe; LL = Lower lobe; OL = Overlapping lesion of lung; AC = Adenocarcinoma; SLC = Squamous lung carcinoma; ASLC = Adenosquamous lung carcinoma; LCLC = Large cell lung carcinoma; SCLC = Small cell lung carcinoma; Surg(pri) = surgical treatments of primary site.
| Insured      | 25275(12.6) | 174934(87.4) | 21607(12.6) | 149701(87.4) |
|-------------|-------------|--------------|-------------|--------------|
| Any Medic aid | 5845(15.5) | 31836(84.5)  | 4945(15.5) | 27004(84.5)  |
| Uninsured   | 1542(20.8)  | 5855(79.2)   | 1419(21.2)  | 5290(78.8)   |
| Unknown     | 670(13.8)   | 4184(86.2)   | 555(13.5)   | 3559(86.5)   |
| Site        |             |              |             |              |
| MB          | 1726(14.5)  | 10146(85.5)  | 1482(14.4)  | 8809(85.6)   |
| UL          | 16803(13.1) | 111712(86.9) | 14390(13.1) | 95484(86.9)  |
| ML          | 1300(12.0)  | 9571(88.0)   | 1117(12.0)  | 8160(88.0)   |
| LL          | 7856(12.2)  | 56663(87.8)  | 6638(12.1)  | 48011(87.9)  |
| OL          | 331(12.2)   | 2391(87.8)   | 278(12.0)   | 2035(88.0)   |
| NOS         | 5316(16.8)  | 26326(83.2)  | 4621(16.7)  | 23055(83.3)  |
| T stage     |             |              |             |              |
| T1          | 3812(7.4)   | 47865(92.6)  | 3185(7.4)   | 39793(92.6)  |
| T2          | 8097(11.9)  | 59812(88.1)  | 6849(11.8)  | 51357(88.2)  |
| T3          | 7148(14.8)  | 41129(85.2)  | 6243(14.9)  | 35732(85.1)  |
| T4          | 9849(17.4)  | 46686(82.6)  | 8392(17.3)  | 40047(82.7)  |
| Unknown     | 4426(17.2)  | 21317(82.8)  | 3857(17.2)  | 18625(82.8)  |
| N stage     |             |              |             |              |
| N0          | 6870(7.2)   | 88264(92.8)  | 5885(7.3)   | 75135(92.7)  |
| N1          | 2724(13.4)  | 17656(86.6)  | 2320(13.2)  | 15228(86.8)  |
| N2          | 14854(16.8) | 73614(83.2)  | 12853(16.8) | 63673(83.2)  |
| N3          | 6612(19.6)  | 27195(80.4)  | 5538(19.4)  | 23033(80.6)  |
| Unknown     | 2272(18.4)  | 10080(81.6)  | 1930(18.5)  | 8485(81.5)   |
| M stage     |             |              |             |              |
| M0          | 2(0.0)      | 120083(100.0)| 0(0.0)      | 102501(100.0)|
| M1          | 33269(25.7) | 95939(74.3)  | 28495(25.6) | 82651(74.4)  |
| Unknown     | 61(7.2)     | 787(92.8)    | 31(7.2)     | 402(92.8)    |

Abbreviations: LC = lung carcinoma; BM = brain metastases; Met = metastases; AI = American Indian/Alaska Native; API = Asian or Pacific Islander; MB = Main bronchus; UL = Upper lobe; ML = Middle lobe; LL = Lower lobe; OL = Overlapping lesion of lung; AC = Adenocarcinoma; SLC = Squamous lung carcinoma; ASLC = Adenosquamous lung carcinoma; LCLC = Large cell lung carcinoma; SCLC = Small cell lung carcinoma; Surg(pri) = surgical treatments of primary site.
| Grade  | Grade I  | 402(3.0) | 13064(97.0) | 347(3.0) | 11084(97.0) |
|--------|----------|----------|-------------|----------|-------------|
| Grade II | 2447(6.3) | 36361(93.7) | 2124(6.4) | 31043(93.6) |
| Grade III | 7564(13.1) | 50167(86.9) | 6537(13.1) | 43491(86.9) |
| Grade IV | 1044(14.8) | 6007(85.2) | 942(15.0) | 5329(85.0) |
| Unknown | 21875(16.4) | 111210(83.6) | 18576(16.4) | 94607(83.6) |
| Bone Met | No | 21576(10.7) | 179603(89.3) | 18546(10.8) | 153784(89.2) |
| | Yes | 11016(23.4) | 36090(76.6) | 9328(23.2) | 30794(76.8) |
| | Unknown | 740(39.9) | 1116(60.1) | 652(40.0) | 976(60.0) |
| Liver Met | No | 25634(11.8) | 191999(88.2) | 21931(11.8) | 164238(88.2) |
| | Yes | 6740(22.2) | 23657(77.8) | 5756(22.1) | 20304(77.9) |
| | Unknown | 958(45.4) | 1153(54.6) | 839(45.3) | 1012(54.7) |
| Lung Met | No | 24048(11.5) | 185211(88.5) | 20557(11.5) | 158472(88.5) |
| | Yes | 8029(21.7) | 29026(78.3) | 6860(21.7) | 24729(78.3) |
| | Unknown | 1255(32.8) | 2572(67.2) | 1109(32.0) | 2353(68.0) |
| Histology | AC | 17409(14.8) | 100406(85.2) | 14710(14.8) | 84956(85.2) |
| | SLC | 3090(5.8) | 50113(94.2) | 2655(5.8) | 42979(94.2) |
| | ASLC | 349(11.2) | 2767(88.8) | 306(11.4) | 2369(88.6) |
| | LCLC | 660(17.8) | 3045(82.2) | 573(17.6) | 2680(82.4) |
| | SCLC | 5503(17.1) | 26675(82.9) | 4727(17.0) | 23040(83.0) |
| | Others | 6321(15.8) | 33803(84.2) | 5555(15.8) | 29530(84.2) |
| Surg(pri) | No | 32382(16.5) | 164148(83.5) | 27712(16.4) | 140781(83.6) |

Abbreviations: LC = lung carcinoma; BM = brain metastases; Met = metastases; AI = American Indian/Alaska Native; API = Asian or Pacific Islander; MB = Main bronchus; UL = Upper lobe; ML = Middle lobe; LL = Lower lobe; OL = Overlapping lesion of lung; AC = Adenocarcinoma; SLC = Squamous lung carcinoma; ASLC = Adenosquamous lung carcinoma; LCLC = Large cell lung carcinoma; SCLC = Small cell lung carcinoma; Surg(pri) = surgical treatments of primary site.
When stratified by different LC histology, the highest prevalence of BM was shown in LCLC patients and the lowest prevalence was indicated in SCL patients. A comparative trend was also indicated in LC patients with distant metastatic disease (Fig. 1B).

**Risk factors for developing BM in LC patients**

As is shown in Table 2, multivariable logistic regression showed Black and API race (vs White), Any Medic aid status and Uninsured status (vs Insured status), higher T stage, N stage, poor differentiated Grade stage, bone metastases, liver metastases, and lung metastases were positively associated with BM, while advanced age and non-AC histology were negatively associated with BM development.

|          | Yes       | Unknown  |
|----------|-----------|----------|
|          | 877(1.7)  | 73(10.8) |
|          | 52056(98.3)| 605(89.2)|
|          | 755(1.7)  | 59(10.3) |
|          | 44261(98.3)| 512(89.7)|

Abbreviations: LC = lung carcinoma; BM = brain metastases; Met = metastases; AI = American Indian/Alaska Native; API = Asian or Pacific Islander; MB = Main bronchus; UL = Upper lobe; ML = Middle lobe; LL = Lower lobe; OL = Overlapping lesion of lung; AC = Adenocarcinoma; SLC = Squamous lung carcinoma; ASLC = Adenosquamous lung carcinoma; LCLC = Large cell lung carcinoma; SCLC = Small cell lung carcinoma; Surg(pri) = surgical treatments of primary site.
Table 2
Univariate and Multivariable Logistic Regression for analyzing the demographic and related clinical characteristics for developing Brain Metastases in patients diagnosed with initial lung carcinoma (diagnosed 2010–2016).

| Subject characteristics | Univariate | Multivariate |
|-------------------------|------------|--------------|
|                         | OR (95%CI) | P-value | OR (95%CI) | P-value |
| Age, years              |            |          |            |          |
| < 41                    | 1.00(Ref)  | 1.00(Ref) | 1.00(Ref)  | 1.00(Ref) |
| 41–50                   | 1.32(1.16–1.49) | < 0.001 | 1.23(1.08–1.40) | 0.002 |
| 51–60                   | 1.16(1.03–1.30) | 0.014 | 1.14(1.01–1.28) | 0.040 |
| 61–70                   | 0.83(0.74–0.93) | 0.001 | 0.89(0.78–1.00) | 0.047 |
| 71–80                   | 0.57(0.51–0.64) | < 0.001 | 0.63(0.56–0.71) | < 0.001 |
| 81–90                   | 0.38(0.34–0.43) | < 0.001 | 0.40(0.35–0.46) | < 0.001 |
| > 90                    | 0.27(0.22–0.32) | < 0.001 | 0.24(0.20–0.29) | < 0.001 |
| Gender                  |            |          |            |          |
| Male                    | 1.00(Ref)  | 1.00(Ref) | NS | NS |
| Female                  | 0.98 (0.96–1.01) | 0.139 | NS | NS |
| Race                    |            |          |            |          |
| White                   | 1.00(Ref)  | 1.00(Ref) | 1.00(Ref)  | 1.00(Ref) |
| Black                   | 1.15(1.11–1.19) | < 0.001 | 1.00(0.96–1.04) | 0.997 |
| AI                      | 1.09(0.93–1.26) | 0.297 | 1.03(0.88–1.21) | 0.718 |
| API                     | 1.31(1.26–1.37) | < 0.001 | 1.19(1.14–1.25) | < 0.001 |
| Unknown                 | 0.81(0.63–1.04) | 0.100 | 0.86(0.67–1.12) | 0.264 |
| Marital status          |            |          |            |          |
| Others                  | 1.00(Ref)  | 1.00(Ref) | 1.00(Ref)  | 1.00(Ref) |
| Married                 | 1.04(1.02–1.07) | < 0.001 | 1.02(1.00–1.05) | 0.080 |
| Unknown                 | 0.89(0.84–0.95) | 0.001 | 0.88(0.82–0.93) | < 0.001 |

Abbreviations: LC = lung carcinoma; BM = brain metastases; Met = metastases; AI = American Indian/Alaska Native; API = Asian or Pacific Islander; MB = Main bronchus; UL = Upper lobe; ML = Middle lobe; LL = Lower lobe; OL = Overlapping lesion of lung; AC = Adenocarcinoma; SLC = Squamous lung carcinoma; ASLC = Adenosquamous lung carcinoma; LCLC = Large cell lung carcinoma; SCLC = Small cell lung carcinoma; Ref = Reference; NS = No significance.
| Subject characteristics | Univariate | Multivariate |
|-------------------------|------------|--------------|
|                         | OR (95%CI) | P-value      | OR (95%CI) | P-value |
| Insurance status        |            |              |            |         |
| Insured                 | 1.00(Ref)  | 1.00(Ref)    | 1.00(Ref)  | 1.00(Ref) |
| Any Medic aid           | 1.27(1.23–1.31) | < 0.001 | 1.04(1.00–1.07) | 0.027 |
| Uninsured               | 1.82(1.72–1.93) | < 0.001 | 1.18(1.11–1.25) | < 0.001 |
| Unknown                 | 1.19(1.02–1.20) | 0.015 | 0.98(0.90–1.07) | 0.696 |
| Site                    |            |              |            |         |
| MB                      | 1.00(Ref)  | 1.00(Ref)    | 1.00(Ref)  | 1.00(Ref) |
| UL                      | 0.88(0.84–0.93) | < 0.001 | 1.15(1.09–1.22) | < 0.001 |
| ML                      | 0.80(0.74–0.86) | < 0.001 | 1.04(0.96–1.13) | 0.299 |
| LL                      | 0.82(0.77–0.86) | < 0.001 | 1.15(1.08–1.22) | < 0.001 |
| OL                      | 0.81(0.72–0.92) | 0.001 | 0.89(0.78–1.01) | 0.078 |
| NOS                     | 1.19(1.12–1.26) | < 0.001 | 1.04(0.98–1.11) | 0.201 |
| T stage                 |            |              |            |         |
| T1                      | 1.00(Ref)  | 1.00(Ref)    | 1.00(Ref)  | 1.00(Ref) |
| T2                      | 1.70(1.63–1.77) | < 0.001 | 1.43(1.37–1.49) | < 0.001 |
| T3                      | 2.18(2.09–2.28) | < 0.001 | 1.48(1.41–1.54) | < 0.001 |
| T4                      | 2.65(2.55–2.76) | < 0.001 | 1.53(1.47–1.60) | < 0.001 |
| Unknown                 | 2.61(2.49–2.73) | < 0.001 | 1.54(1.46–1.63) | < 0.001 |
| N stage                 |            |              |            |         |
| N0                      | 1.00(Ref)  | 1.00(Ref)    | 1.00(Ref)  | 1.00(Ref) |
| N1                      | 1.98(1.89–2.08) | < 0.001 | 1.54(1.47–1.62) | < 0.001 |
| N2                      | 2.59(2.52–2.67) | < 0.001 | 1.68(1.62–1.73) | < 0.001 |
| N3                      | 3.12(3.01–3.24) | < 0.001 | 1.69(1.62–1.76) | < 0.001 |
| Unknown                 | 2.90(2.75–3.05) | < 0.001 | 1.66(1.56–1.76) | < 0.001 |

Abbreviations: LC = lung carcinoma; BM = brain metastases; Met = metastases; AI = American Indian/Alaska Native; API = Asian or Pacific Islander; MB = Main bronchus; UL = Upper lobe; ML = Middle lobe; LL = Lower lobe; OL = Overlapping lesion of lung; AC = Adenocarcinoma; SLC = Squamous lung carcinoma; ASLC = Adenosquamous lung carcinoma; LCLC = Large cell lung carcinoma; SCLC = Small cell lung carcinoma; Ref = Reference; NS = No significance.
| Subject characteristics | Univariate |          | Multivariate |          |
|-------------------------|------------|----------|--------------|----------|
|                         | OR (95%CI) | P-value  | OR (95%CI)   | P-value  |
| Grade                   |            |          |              |          |
| Grade I                 | 1.00(Ref)  | 1.00(Ref)| 1.00(Ref)    | 1.00(Ref)|
| Grade II                | 2.19(1.96–2.44) | < 0.001 | 2.19(1.96–2.44) | < 0.001 |
| Grade III               | 4.90(4.42–5.43) | < 0.001 | 3.75(3.38–4.17) | < 0.001 |
| Grade IV                | 5.65(5.01–6.36) | < 0.001 | 3.58(3.16–4.05) | < 0.001 |
| Unknown                 | 6.39(5.78–7.07) | < 0.001 | 4.23(3.82–4.69) | < 0.001 |
| Bone Met                |            |          |              |          |
| No                      | 1.00(Ref)  | 1.00(Ref)| 1.00(Ref)    | 1.00(Ref)|
| Yes                     | 2.54(2.48–2.61) | < 0.001 | 1.58(1.54–1.62) | < 0.001 |
| Unknown                 | 5.52(5.03–6.06) | < 0.001 | 2.22(1.99–2.48) | < 0.001 |
| Liver Met               |            |          |              |          |
| No                      | 1.00(Ref)  | 1.00(Ref)| 1.00(Ref)    | 1.00(Ref)|
| Yes                     | 2.13(2.07–2.20) | < 0.001 | 1.26(1.22–1.30) | < 0.001 |
| Unknown                 | 6.22(5.71–6.79) | < 0.001 | 2.72(2.45–3.01) | < 0.001 |
| Lung Met                |            |          |              |          |
| No                      | 1.00(Ref)  | 1.00(Ref)| 1.00(Ref)    | 1.00(Ref)|
| Yes                     | 2.13(2.07–2.19) | < 0.001 | 1.45(1.40–1.49) | < 0.001 |
| Unknown                 | 3.76(3.51–4.03) | < 0.001 | 1.83(1.69–1.98) | < 0.001 |
| Histology               |            |          |              |          |
| AC                      | 1.00(Ref)  | 1.00(Ref)| 1.00(Ref)    | 1.00(Ref)|
| SLC                     | 0.36(0.34–0.37) | < 0.001 | 0.39(0.37–0.40) | < 0.001 |
| ASLC                    | 0.73(0.65–0.81) | < 0.001 | 0.71(0.64–0.80) | < 0.001 |
| LCLC                    | 1.25(1.15–1.36) | < 0.001 | 1.04(0.95–1.14) | 0.409   |
| SCLC                    | 1.19(1.15–1.23) | < 0.001 | 0.81(0.78–0.84) | < 0.001 |

Abbreviations: LC = lung carcinoma; BM = brain metastases; Met = metastases; AI = American Indian/Alaska Native; API = Asian or Pacific Islander; MB = Main bronchus; UL = Upper lobe; ML = Middle lobe; LL = Lower lobe; OL = Overlapping lesion of lung; AC = Adenocarcinoma; SLC = Squamous lung carcinoma; ASLC = Adenosquamous lung carcinoma; LCLC = Large cell lung carcinoma; SCLC = Small cell lung carcinoma; Ref = Reference; NS = No significance.
### Subject characteristics

| Subject characteristics | Univariate | | Multivariate |
|-------------------------|------------|--|--------------|
|                        | OR (95%CI) | P-value | OR (95%CI) | P-value |
| Others                 | 1.08(1.05–1.11) | < 0.001 | 0.96(0.93–0.99) | 0.015 |

**Abbreviations**: LC = lung carcinoma; BM = brain metastases; Met = metastases; AI = American Indian/Alaska Native; API = Asian or Pacific Islander; MB = Main bronchus; UL = Upper lobe; ML = Middle lobe; LL = Lower lobe; OL = Overlapping lesion of lung; AC = Adenocarcinoma; SLC = Squamous lung carcinoma; ASLC = Adenosquamous lung carcinoma; LCLC = Large cell lung carcinoma; SCLC = Small cell lung carcinoma; Ref = Reference; NS = No significance.

---

### Construction of nomogram for predicting BM development and performance assessment

Basing on the aforementioned risk factors for BM, a nomogram was constructed to accurately predict the probability for development BM (Fig. 2A). The AUC of the ROC curve 73.1% (95% CI: 72.6–73.6%) which indicated good discrimination (Fig. 2B). The stability of this prediction model was validated by random splitting method, the AUC values of construction and validation model were 72.5% (95% CI: 72.0–72.9%) and 72.6% (95% CI: 72.3–72.9%), respectively, with no significant difference (D = -0.43; P = 0.67) (Fig. 2B). The calibration curve revealed good agreement between the predicted and observed probabilities for BM occurrence, which indicated good calibration ability of the nomogram (Fig. 2C).

### Survival analysis

For the 28,526 patients who developed BM at initial diagnosis of LC, the 1-year, 3-year and 5-year overall survival were 21%, 5% and 3%, respectively, with the median overall survival (OS) was 4 months (95% CI: 3.91–4.09 months).

Multivariate Cox regression analysis showed, older age, insured status, higher T stage, N stage, poor differentiated tumor grade; tumor located in the and the presence of lung, liver, or bone metastases were positively associated overall death, while female gender, API race, married status, UL site, and surgery on primary site were negatively associated with overall mortality (Table 3).
### Table 3

Multivariable Cox Regression for analyzing the prognosis factors among primary lung carcinoma with brain metastases (diagnosed 2010–2015).

| Subject characteristics | Patients’ No. of LC with BM | Median Survival (IQR), mo | Univariate | Multivariate |
|-------------------------|-----------------------------|---------------------------|------------|--------------|
|                         | Overall                     | Dead (N, %)               |HR (95%CI) | P-value      |HR (95%CI) | P-value |
| Age, years              |                             |                           |            |              |            |         |
| <41                     | 303                         | 231(76.2)                 | 13.00(10.29–15.72) | 1.00 (Ref) | 1.00(Ref) | 1.00 |
| 41–50                   | 2011                        | 1768(87.9)                | 8.00(7.48–8.52)   | 1.36(1.18–1.55) | < 0.001 | 1.36(1.18–1.56) | < 0.001 |
| 51–60                   | 7586                        | 6885(90.8)                | 6.00(5.77–6.24)   | 1.66(1.46–1.89) | < 0.001 | 1.67(1.47–1.91) | < 0.001 |
| 61–70                   | 9848                        | 9158(93.0)                | 4.00(3.85–4.15)   | 1.92(1.69–2.19) | < 0.001 | 1.98(1.74–2.26) | < 0.001 |
| 71–80                   | 6479                        | 6191(95.6)                | 3.00(2.90–3.11)   | 2.44(2.14–2.78) | < 0.001 | 2.59(2.27–2.96) | < 0.001 |
| 81–90                   | 2145                        | 2094(97.6)                | 2.00(1.90–2.11)   | 3.25(2.83–3.72) | < 0.001 | 3.37(2.94–3.87) | < 0.001 |
| >90                     | 154                         | 153(99.4)                 | 1.00(0.70–1.30)   | 4.88(3.98–5.99) | < 0.001 | 4.52(3.68–5.55) | < 0.001 |
| Sex                     |                             |                           |            |              |            |         |
| Male                    | 14907                       | 14024(94.1)               | 3.00(2.89–3.11)   | 1.00(Ref)  | 1.00(Ref) | 1.00 |
| Female                  | 13619                       | 12456(91.5)               | 4.00(3.83–4.17)   | 0.84(0.82–0.86) | < 0.001 | 0.84(0.82–0.86) | < 0.001 |
| Race                    |                             |                           |            |              |            |         |
| White                   | 22400                       | 20931(93.4)               | 4.00(3.90–4.10)   | 1.00(Ref)  | 1.00(Ref) | 1.00 |
| Black                   | 3619                        | 3406(94.1)                | 4.00(3.77–4.23)   | 1.00(0.97–1.04) | 0.816 | 1.00(0.97–1.04) | 0.889 |
| AI                      | 159                         | 154(96.9)                 | 3.00(2.05–3.95)   | 1.17(1.00–1.37) | 0.055 | 1.16(0.99–1.36) | 0.067 |

Abbreviations: LC = lung carcinoma; BM = brain metastases; Met = metastases; AI = American Indian/Alaska Native; API = Asian or Pacific Islander; MB = Main bronchus; UL = Upper lobe; ML = Middle lobe; LL = Lower lobe; OL = Overlapping lesion of lung; AC = Adenocarcinoma; SLC = Squamous lung carcinoma; ASLC = Adenosquamous lung carcinoma; LCLC = Large cell lung carcinoma; SCLC = Small cell lung carcinoma; Surg(pri) = surgical treatments of primary site.
| Subject characteristics       | Patients' No. of LC with BM | Median Survival (IQR), mo | Univariate | Multivariate |
|-------------------------------|----------------------------|--------------------------|------------|--------------|
|                              | Overall  | Dead (N, %) | Overall  | Dead (N, %) | HR (95%CI) | P-value  | HR (95%CI) | P-value |
| API                           | 2296     | 1951(85.0) | 7.00(6.30–7.70) | 0.68(0.65–0.71) | < 0.001 | 0.69(0.65–0.72) | < 0.001 |
| Unknown                       | 52       | 38(73.1)   | 7.00(3.94–10.06) | 0.76(0.55–1.04) | 0.090  | 0.73(0.53–1.00) | 0.053  |
| Marital status                |          |            |            |              |        |                      |        |
| Others                        | 13123    | 12372(94.3) | 3.00(2.89–3.11) | 1.00(Ref) | 1.00  | 1.00(Ref) | 1.00  |
| Married                       | 14238    | 13014(91.4) | 5.00(4.84–5.16) | 0.82(0.80–0.84) | < 0.001 | 0.86(0.83–0.88) | < 0.001 |
| Unknown                       | 1165     | 1094(93.9)  | 4.00(3.59–4.41) | 0.96(0.90–1.02) | 0.196  | 0.95(0.89–1.01) | 0.091  |
| Insurance status              |          |            |            |              |        |                      |        |
| Insured                       | 21607    | 20018(92.7) | 4.00(3.89–4.11) | 1.00(Ref) | 1.00  | 1.00(Ref) | 1.00  |
| Any Medic aid                 | 4945     | 4619(93.4)  | 4.00(3.80–4.20) | 1.07(1.04–1.11) | < 0.001 | 1.17(1.13–1.21) | < 0.001 |
| Uninsured                     | 1419     | 1322(93.2)  | 4.00(3.66–4.34) | 1.09(1.04–1.16) | 0.001  | 1.27(1.20–1.34) | < 0.001 |
| Unknown                       | 555      | 521(93.9)   | 3.00(2.61–3.39) | 1.21(1.11–1.32) | < 0.001 | 1.13(1.03–1.23) | 0.010  |
| Site                          |          |            |            |              |        |                      |        |
| MB                            | 1482     | 1415(95.5)  | 3.00(2.64–3.36) | 1.00(Ref) | 1.00  | 1.00(Ref) | 1.00  |
| UL                            | 14390    | 13307(92.5) | 4.00(3.85–4.15) | 0.83(0.79–0.88) | < 0.001 | 0.90(0.86–0.96) | < 0.001 |
| ML                            | 1117     | 1025(91.8)  | 4.00(3.54–4.46) | 0.81(0.75–0.88) | < 0.001 | 0.89(0.82–0.97) | 0.006  |
| LL                            | 6638     | 6140(92.5)  | 4.00(3.82–4.19) | 0.86(0.82–0.92) | < 0.001 | 0.94(0.88–0.99) | 0.032  |

Abbreviations: LC = lung carcinoma; BM = brain metastases; Met = metastases; AI = American Indian/Alaska Native; API = Asian or Pacific Islander; MB = Main bronchus; UL = Upper lobe; ML = Middle lobe; LL = Lower lobe; OL = Overlapping lesion of lung; AC = Adenocarcinoma; SLC = Squamous lung carcinoma; ASLC = Adenosquamous lung carcinoma; LCLC = Large cell lung carcinoma; SCLC = Small cell lung carcinoma; Surg(pri) = surgical treatments of primary site.
| Subject characteristics | Patients’ No. of LC with BM | Median Survival (IQR), mo | Univariate | Multivariate |
|-------------------------|----------------------------|--------------------------|------------|-------------|
|                         | Overall                    | Dead (N, %)               | HR (95%CI) | P-value     | HR (95%CI) | P-value |
| OL                      | 278                        | 263(94.6)                | 3.00(1.90–4.11) | 0.94(0.82–1.07) | 0.337 | 1.01(0.88–1.15) | 0.927 |
| NOS                     | 4621                       | 4330(93.7)               | 3.00(2.83–3.17) | 0.97(0.92–1.03) | 0.348 | 0.97(0.91–1.03) | 0.339 |
| T stage                 |                            |                          |            |             |            |         |
| T1                      | 3185                       | 2818(88.5)               | 6.00(5.59–6.41) | 1.00(Ref) | 1.00 | 1.00(Ref) | 1.00 |
| T2                      | 6849                       | 6298(92.0)               | 4.00(3.78–4.22) | 1.21(1.15–1.26) | < 0.001 | 1.17(1.12–1.22) | < 0.001 |
| T3                      | 6243                       | 5822(93.3)               | 4.00(3.84–4.16) | 1.34(1.28–1.40) | < 0.001 | 1.25(1.19–1.31) | < 0.001 |
| T4                      | 8392                       | 7913(94.3)               | 3.00(2.85–3.15) | 1.40(1.35–1.47) | < 0.001 | 1.29(1.23–1.34) | < 0.001 |
| Unknown                 | 3857                       | 3629(94.1)               | 3.00(2.81–3.19) | 1.49(1.42–1.56) | < 0.001 | 1.28(1.22–1.35) | < 0.001 |
| N stage                 |                            |                          |            |             |            |         |
| N0                      | 5885                       | 5304(90.1)               | 4.00(3.75–4.25) | 1.00(Ref) | 1.00 | 1.00(Ref) | 1.00 |
| N1                      | 2320                       | 2143(92.4)               | 4.00(3.64–4.36) | 1.06(1.00–1.11) | 0.033 | 1.02(0.97–1.07) | 0.569 |
| N2                      | 12853                      | 12042(93.7)              | 4.00(3.87–4.13) | 1.16(1.12–1.20) | < 0.001 | 1.09(1.06–1.13) | < 0.001 |
| N3                      | 5538                       | 5162(93.2)               | 4.00(3.80–4.20) | 1.15(1.10–1.19) | < 0.001 | 1.09(1.05–1.14) | < 0.001 |
| Unknown                 | 1930                       | 1829(94.8)               | 3.00(2.78–3.22) | 1.36(1.29–1.44) | < 0.001 | 1.09(1.03–1.16) | 0.003 |
| Grade                   |                            |                          |            |             |            |         |
| Grade I                 | 347                        | 306(88.2)                | 7.00(5.18–8.83) | 1.00(Ref) | 1.00 | 1.00(Ref) | 1.00 |

Abbreviations: LC = lung carcinoma; BM = brain metastases; Met = metastases; AI = American Indian/Alaska Native; API = Asian or Pacific Islander; MB = Main bronchus; UL = Upper lobe; ML = Middle lobe; LL = Lower lobe; OL = Overlapping lesion of lung; AC = Adenocarcinoma; SLC = Squamous lung carcinoma; ASLC = Adenosquamous lung carcinoma; LCLC = Large cell lung carcinoma; SCLC = Small cell lung carcinoma; Surg(pri) = surgical treatments of primary site.
| Subject characteristics | Patients' No. of LC with BM | Median Survival (IQR), mo | Univariate | Multivariate |
|--------------------------|-----------------------------|---------------------------|------------|-------------|
|                          | Overall Dead (N, %)         |                           | HR (95%CI) | P-value     | HR (95%CI) | P-value |
| Grade II                 | 2124 1876 (88.3)            | 6.00 (5.49–6.51)          | 1.04 (0.92–1.17) | 0.552       | 1.01 (0.89–1.14) | 0.904   |
| Grade III                | 6537 6115 (93.5)            | 4.00 (3.85–4.15)          | 1.42 (1.26–1.59) | < 0.001     | 1.22 (1.09–1.37) | 0.001   |
| Grade IV                 | 942 907 (96.3)              | 4.00 (3.52–4.49)          | 1.48 (1.30–1.68) | < 0.001     | 1.21 (1.06–1.38) | 0.004   |
| Unknown                  | 18576 17276 (93.0)          | 4.00 (3.88–4.12)          | 1.42 (1.27–1.59) | < 0.001     | 1.15 (1.03–1.29) | 0.017   |
| Bone Met                 |                             |                           |            |             |            |         |
| No                       | 18546 17085 (92.1)          | 4.00 (3.87–4.13)          | 1.00 (Ref) | 1.00        | 1.00 (Ref) | 1.00    |
| Yes                      | 9328 8772 (94.0)            | 3.00 (2.87–3.13)          | 1.16 (1.13–1.19) | < 0.001     | 1.13 (1.10–1.17) | < 0.001 |
| Unknown                  | 652 623 (95.6)              | 3.00 (2.55–3.45)          | 1.28 (1.18–1.38) | < 0.001     | 1.03 (0.94–1.14) | 0.504   |
| Liver Met                |                             |                           |            |             |            |         |
| No                       | 21931 20117 (91.7)          | 4.00 (3.88–4.12)          | 1.00 (Ref) | 1.00        | 1.00 (Ref) | 1.00    |
| Yes                      | 5756 5556 (96.5)            | 2.00 (1.88–2.13)          | 1.45 (1.41–1.49) | < 0.001     | 1.35 (1.31–1.39) | < 0.001 |
| Unknown                  | 839 807 (96.2)              | 3.00 (2.67–3.33)          | 1.30 (1.22–1.40) | < 0.001     | 1.13 (1.04–1.24) | 0.007   |
| Lung Met                 |                             |                           |            |             |            |         |
| No                       | 20557 18943 (92.2)          | 4.00 (3.88–4.12)          | 1.00 (Ref) | 1.00        | 1.00 (Ref) | 1.00    |
| Yes                      | 6860 6478 (94.4)            | 3.00 (2.84–3.16)          | 1.17 (1.14–1.20) | < 0.001     | 1.07 (1.04–1.11) | < 0.001 |
| Unknown                  | 1109 1059 (95.5)            | 3.00 (2.72–3.28)          | 1.24 (1.16–1.32) | < 0.001     | 1.03 (0.95–1.11) | 0.481   |
| Histology                |                             |                           |            |             |            |         |

Abbreviations: LC = lung carcinoma; BM = brain metastases; Met = metastases; AI = American Indian/Alaska Native; API = Asian or Pacific Islander; MB = Main bronchus; UL = Upper lobe; ML = Middle lobe; LL = Lower lobe; OL = Overlapping lesion of lung; AC = Adenocarcinoma; SLC = Squamous lung carcinoma; ASLC = Adenosquamous lung carcinoma; LCLC = Large cell lung carcinoma; SCLC = Small cell lung carcinoma; Surg(pri) = surgical treatments of primary site.
| Subject characteristics | Patients’ No. of LC with BM | Median Survival (IQR), mo | Univariate | Multivariate |
|--------------------------|----------------------------|--------------------------|------------|--------------|
|                          | Overall Dead (N, %)        |                          | HR (95%CI) | P-value      | HR (95%CI) | P-value |
| AC                       | 14710 13167(89.5)          | 5.00(4.85–5.15)          | 1.00(Ref)  | 1.00         | 1.00(Ref)  | 1.00    |
| SLC                      | 2655 2563(96.5)            | 3.00(2.84–3.16)          | 1.54(1.47–1.60) | < 0.001 | 1.40(1.34–1.46) | < 0.001 |
| ASLC                     | 306 286(93.5)              | 4.00(3.21–4.79)          | 1.18(1.05–1.33) | 0.005 | 1.17(1.04–1.32) | 0.008   |
| LCLC                     | 573 539(94.1)              | 4.00(3.43–4.57)          | 1.26(1.16–1.37) | < 0.001 | 1.23(1.13–1.34) | < 0.001 |
| SCLC                     | 4727 4559(96.5)            | 5.00(4.72–5.29)          | 1.28(1.23–1.32) | < 0.001 | 1.11(1.07–1.15) | < 0.001 |
| Others                   | 5555 5366(96.6)            | 2.00(1.90–2.10)          | 1.76(1.71–1.82) | < 0.001 | 1.59(1.54–1.64) | < 0.001 |
| Surg(pri)                |                            |                          |            |              |            |         |
| No                       | 27712 25867(93.3)          | 4.00(3.91–4.091)         | 1.00(Ref)  | 1.00         | 1.00(Ref)  | 1.00    |
| Yes                      | 755 560(74.2)              | 13.00(11.41–14.59)       | 0.89(0.68–1.17) | 0.415 | 0.55(0.50–0.60) | < 0.001 |
| Unknown                  | 59 53(89.8)                | 2.00(1.20–2.80)          | 0.39(0.30–0.52) | < 0.001 | 0.86(0.65–1.12) | 0.261   |

Abbreviations: LC = lung carcinoma; BM = brain metastases; Met = metastases; AI = American Indian/Alaska Native; API = Asian or Pacific Islander; MB = Main bronchus; UL = Upper lobe; ML = Middle lobe; LL = Lower lobe; OL = Overlapping lesion of lung; AC = Adenocarcinoma; SLC = Squamous lung carcinoma; ASLC = Adenosquamous lung carcinoma; LCLC = Large cell lung carcinoma; SCLC = Small cell lung carcinoma; Surg(pri) = surgical treatments of primary site.

**Construction of nomogram for predicting prognosis and performance assessment**

The survival predicted nomogram was constructed basing on the prognostic factors for predicting the 1-year, 3-year and 5-year overall survival of LC patients with BM (Fig. 3A). The overall C-index was 0.88 (95% CI 0.87–0.89), indicating that the nomogram has a satisfactory performance of distinguishing people who have terminal events from those who do not.

The calibration curve has good performance for predicting overall survival in 1-year (Fig. 3B), 3-year (Fig. 3C) and 5-year (Fig. 3D). Additionally, the LC patients who were diagnosed in 2016 were also selected to validate
the performance of the nomogram and the calibration curve also suggested good agreement between predicted and actual probability curve (Fig. 3E).

Discussion

The present study comprehensively described the prevalence of BM in patients with initially diagnosed LC in the SEER cohort, then the risk and prognostic factors associated with BM were detected. Then we conducted two nomograms to predict the development and prognosis of LC patients, respectively. The results showed the prevalence of BM in LC patients was 16.25%, which was consistent with previous studies that ranged from 10.4%-19.9% [3, 16]. In accordance with the consequence of risk factors analysis, we specified that LC patients are more likely to be diagnosed with BM when have the following features, age at 41–60 years, API race, Any Medic aid status and Uninsured status, tumor located in Upper lobe and Lower lobe, higher T stage, more lymphatic metastasis (N), higher Grade stage, the presence of metastases out of brain and AC histological subtype. Part of the above risk factors also has been determined previously in studies [17–20]. To the best of our knowledge, this study firstly put insight into and confirmed tumor site as one of the risk factors associated with the development of BM. It's worth noting that a few works previously determined that a higher incidence of BM in women with NSCLC than in men [3, 21, 22]. No significant differences in the BM incidence of LC with no specific histological subtype according to sex were observed in our study. More studies are warranted to illuminate the result and potential cause.

A large cohort containing 28,526 LC patients with BM who were followed up for more than 1 year was utilized for survival analysis, results suggested that the median overall survival is 4 months, and a similar consequence has been reported before [4, 22]. It has been widely discussed in many pieces of research that older age, poor tumor grade, presence of extracranial metastases, non-AC histological subtype and the absence of definitive therapy were the negative relevant predictors to the prognosis of LC patients with BM [23–26]. In the present study, we found that patients have the following features also shown less favorable outcomes: male; uninsured status; higher T stage; lymph node involvement; tumor located in the main bronchus. No previous studies reported that API race patients were to get better survival than other races, but which was determined in our study. Further researches are needed to reveal the potential cause.

National Comprehensive Cancer Network (NCCN) guidelines do not recommend BM screening for patients who are asymptomatic and clinical stage I [9], despite the brain imaging has been increasingly extensively applied to newly diagnosed LC patients, there are still a large number of missed diagnoses of BM in early-stage LC or asymptomatic patients [27]. To get timely BM screening for the suitable population, an easy operating prediction nomogram was constructed in our data set. A quantified metastatic risk could be generated utilizing the constructed nomogram based on established risk factors, including patient’s age, race, insurance status, primary tumor site, primary tumor, N stage, tumor differentiation grade, the presence or absence of lung metastases, liver metastases, or bone metastases, histology. The performance of the nomogram was evaluated from two aspects; it is qualified for prediction the risk of BM development. Furthermore, the results of internal validation indicated the excellent stability of the accurate prediction effect of this model.
Many parameters influence the choice of treatment for patients with BM of LC [23], it is vital to determine more effective forms of treatment to achieve a longer survival and better quality of life [25]. The nomogram for predicting OS was built based on factors determined by multivariate Cox regression analysis, the C-index was 0.88, with the addition of the calibration curve were the proof of prediction capacity of the model. Some previous studies have established different scoring systems to evaluate the prognosis of LC patients with BM: Recursive Partitioning Analysis (RPA), Gaspar L, et al [24]; Diagnosis-Specific Graded Prognostic Assessment (DS-GPA), Sperduto PW, et al [28]; Graded Prognostic Assessment for Lung Cancer Using Molecular Markers (Lung-molGPA), Paul W. Sperduto, et al [19]. From which we have learned that other factors besides our model like the number of brain metastases, gene alterations and treatment for brain metastases also contribute to the influence of patients’ survival.

Our study had several limitations came from the SEER database itself. First, the actual BM rate in LC patients might be undervalued due to non-recording in asymptomatic cases and missed information related to BM development later during the disease course. Second, data about comorbidities, smoking history and Molecular biological characteristics of patients are not available, those were supposed to be the underlying risk and prognosis factors. Third, as a result of data access limitations, the specific treatment program of the study subjects could not be obtained. Finally, the follow-up time of the external validation group was short. These limitations may affect the prediction accuracy of the model to some extent. On the other hand, notwithstanding the calibration and discrimination ability of the prediction models reached a high level, suggesting that the prediction model may be applied to a more extensive and heterogeneous population, considering that the deficiencies of the external validation set, the research results still need to be further verified.

**Conclusion**

In conclusion, BM was highly developed in LC patients with a prevalence of 16.25%, which was fluctuated with age. The prognosis of LC patients with BM was relatively poor with a median OS of 4 months. The nomograms have good performance and stability for predicting the BM development and prognosis, which may help the clinicians to identify the high-risk patients and tailor targeted treatment regimens for LC patients with BM to improve the prognosis.

**Abbreviations**

BM  
Brain metastasis  
LC  
Lung cancer  
SEER  
Surveillance, Epidemiology, and End Results Program  
ROC curve  
receiver operating characteristics curve  
AI
Declarations

Ethics statement

Cancer information in the SEER database is continually reported in every state of the United States and retrieved with no need for informed patient consent. Our study complied with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and analysed during the current study are available in the SEER repository. SEER*Stat Software version 8.3.4 (https://seer.cancer.gov/seerstat/) (Information Management Service, Inc. Calverton, MD, USA) was used to generate the case listing.

Competing interests

The authors declare no potential conflicts of interest.

Funding

Not applicable.

Authors’ contributions
All authors have had access to the data and all drafts of the manuscript. Specific contributions are as follows: study design: HC, CZ, GL, JT; data collection: CZ, YB, GL, XL; data management and analysis: CZ, PS; manuscript drafting: PC, CZ; manuscript review: all. All authors read and approved the final manuscript.

Acknowledgments

We thank Xin Wang for his contribution on collecting data and editing the language of the manuscript. This study used the linked SEER-Medicare database. The interpretation and reporting of these data are the sole responsibility of the authors. The authors acknowledge the efforts of the National Cancer Institute; the Office of Research, Development and Information, CMS; Information Management Services (IMS), Inc.; and the Surveillance, Epidemiology, and End Results (SEER) Program tumor registries in the creation of the SEER-Medicare database.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel R L, Torre L A, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin, 2018. 68(6): p. 394-424.

2. Guo X, Zhang C, Ma W, Tian F, Xu G, Han X, et al. Patterns of bone metastases in newly diagnosed colorectal cancer: a real-world analysis in the SEER database. Int J Colorectal Dis, 2019. 34(3): p. 533-543.

3. Barnholtz-Sloan JS, Sloan A, Davis FG, Vigneau FD, Lai P, Sawaya RE. Incidence proportions of brain metastases in patients diagnosed (1973 to 2001) in the Metropolitan Detroit Cancer Surveillance System. J Clin Oncol, 2004. 22(14): p. 2865-72.

4. Cagney DN, Martin AM, Catalano PJ, Redig AJ, Lin NU, Lee EQ, et al. Incidence and prognosis of patients with brain metastases at diagnosis of systemic malignancy: a population-based study. Neuro Oncol, 2017. 19(11): p. 1511-1521.

5. Nayak L, Lee EQ, Wen PY. Epidemiology of Brain Metastases. Current Oncology Reports, 2011. 14(1): p. 48-54.

6. Nam MW, Kim JM, Cheong JH, Ryu JI, Han MH. Paradoxical relationship between body mass index and bone mineral density in patients with non-small cell lung cancer with brain metastasis. PLoS One, 2019. 14(6): p. e0218825.

7. Pedrosa R, Mustafa D AM, Aerts Jgjv, Kros JM. Potential Molecular Signatures Predictive of Lung Cancer Brain Metastasis. Front Oncol, 2018. 8: p. 159.

8. Ascha MS, Ostrom QT, Wright J, Kumthekar P, Bordeaux JS, Sloan AE, et al. Lifetime Occurrence of Brain Metastases Arising from Lung, Breast, and Skin Cancers in the Elderly: A SEER-Medicare Study. Cancer Epidemiol Biomarkers Prev, 2019. 28(5): p. 917-925.

9. Rueda-Clausen CF, Ogunleye AA, Sharma AM. Health Benefits of Long-Term Weight-Loss Maintenance. Annu Rev Nutr, 2015. 35: p. 475-516.

10. Gong L, Wang QI, Zhao L, Yuan Z, Li R, Wang P. Factors affecting the risk of brain metastasis in small cell lung cancer with surgery: is prophylactic cranial irradiation necessary for stage I-II disease? Int J
11. Yu X, Fan Y. Real-World Data on Prognostic Factors for Overall Survival in EGFR-Mutant Non-Small-Cell Lung Cancer Patients with Brain Metastases. J Cancer, 2019. 10(15): p. 3486-3493.

12. Harada H, Asakura Hirofumi, Ogawa Hirofumi, Mori Keita, Takahashi Toshiaki, Nakasu Yoko, et al. Prognostic factors in patients with brain metastasis from non-small cell lung cancer treated with whole-brain radiotherapy. Journal of Cancer Research and Therapeutics, 2016. 12(1): p. 267-270.

13. Zheng Y, Wang L, Zhao W, Dou Y, Lv W, Yang H, et al. Risk factors for brain metastasis in patients with small cell lung cancer without prophylactic cranial irradiation. Strahlenther Onkol, 2018. 194(12): p. 1152-1162.

14. Kim YJ, Kim JS, Kim IA. Molecular subtype predicts incidence and prognosis of brain metastasis from breast cancer in SEER database. J Cancer Res Clin Oncol, 2018. 144(9): p. 1803-1816.

15. Zhang F, Zheng W, Ying L, Wu J, Wu S, Ma S, Bradley J, Devarakonda S, Du L, et al. A Nomogram to Predict Brain Metastases of Resected Non-Small Cell Lung Cancer Patients. Ann Surg Oncol, 2016. 23(9): p. 3033-9.

16. Waqar SN, Samson PP, Robinson CG, et al., Non-small-cell Lung Cancer With Brain Metastasis at Presentation. Clin Lung Cancer, 2018. 19(4): p. e373-e379.

17. Hubbs JL, DuL, Hollis D, Chino JP, Saynak M, Kelsey CR. Factors associated with the development of brain metastases: analysis of 975 patients with early stage nonsmall cell lung cancer. Cancer, 2010. 116(21): p. 5038-46.

18. Lim JH, Um SW. The risk factors for brain metastases in patients with non-small cell lung cancer. Ann Transl Med, 2018. 6(Suppl 1): p. S66.

19. Sperduto PW, Yang TJ, Beal K, Pan H, Brown PD, Bangdiwala A, et al. Estimating Survival in Patients With Lung Cancer and Brain Metastases: An Update of the Graded Prognostic Assessment for Lung Cancer Using Molecular Markers (Lung-molGPA). JAMA Oncol, 2017. 3(6): p. 827-831.

20. Yawn BP, Wollan PC, Schroeder C, Gazvuola L, Mehta M. Temporal and gender-related trends in brain metastases from lung and breast cancer. Minn Med, 2003. 86(12): p. 32-7.

21. Goncalves PH, Peterson SL, Vigneau FD, Shore RD, Quashie WO, Islam K, et al. Risk of brain metastases in patients with nonmetastatic lung cancer: Analysis of the Metropolitan Detroit Surveillance, Epidemiology, and End Results (SEER) data. Cancer, 2016. 122(12): p. 1921-7.

22. Riihimaki M, Hemminki A, Fallah M, Thomsen H, Sundquist K, Sundquist J, et al. Metastatic sites and survival in lung cancer. Lung Cancer, 2014. 86(1): p. 78-84.

23. Jolly K, Chambers R. Improving outcomes for patients with obesity. Practitioner, 2014. 258(1773): p. 29-31, 3.

24. Bernhardt D, Adeberg S, Bozorgmehr F, Opfermann N, Horner-Rieber J, König L, et al. Outcome and prognostic factors in single brain metastases from small-cell lung cancer. Strahlenther Onkol, 2018. 194(2): p. 98-106.

25. Andratschke N, Kraft J, Nieder C, Tay R, Califano R, Soffietti R, et al. Optimal management of brain metastases in oncogenic-driven non-small cell lung cancer (NSCLC). Lung Cancer, 2019. 129: p. 63-71.
26. Zindler JD, Jochems A, Lagerwaard FJ, Beumer R, Troost EGC, Eekers DBP, et al. Individualized early death and long-term survival prediction after stereotactic radiosurgery for brain metastases of non-small cell lung cancer: Two externally validated nomograms. Radiother Oncol, 2017. 123(2): p. 189-194.

27. Ando T, Kage H, Saito M, Amano Y, Goto Y, Nakajima J, et al. Early stage non-small cell lung cancer patients need brain imaging regardless of symptoms. Int J Clin Oncol, 2018. 23(4): p. 641-646.

28. Sperduto PW, Kased N, Roberge D, Xu Z, Shanley R, Luo X, et al. Summary report on the graded prognostic assessment: an accurate and facile diagnosis-specific tool to estimate survival for patients with brain metastases. J Clin Oncol, 2012. 30(4): p. 419-25.

Supplemental Figure

Supplementary Figure 1: The flow-chart for the population selection.

Figures
Figure 2

The prevalence of brain metastasis in lung cancer patients with metastasis or not. A: stratified by age; B: stratified by histology.
Figure 4

The nomogram for predicting brain metastasis in lung cancer (A), ROC curve and randomly splitting method for evaluating the performance and stability of the nomogram (B) and the calibration curve for estimating the predictive accuracy of the nomogram (C).
Figure 6

The nomogram for predicting prognosis of lung cancer patients with brain metastasis (A) and the calibration curve for evaluating the performance in predicting 1-year (B), 3-year (C), 5-year (D) and external performance in 1-year (E) survival probability.

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

- Supplementaryfigure1.tif
- Supplementaryfigure1.tif