

A risk score-matched model analysis of 4,452 patients with small cell lung cancer: who will benefit from operative treatment

Shuzhen Tan

Guangdong Medical University

Ying Kong

Hubei No.3 People's Hospital of Jiangnan University

Shuilong Leng

Guangzhou Medical University

Xiao Zhu (✉ xzhu@gdmu.edu.cn)

Guangdong Medical University <https://orcid.org/0000-0002-1737-3386>

Research article

Keywords: small-cell lung cancer, SEER, Nomogram, Cox regression model, ROC curve, training cohort, data cleaning

Posted Date: June 10th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-32351/v1>

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

[Read Full License](#)

Abstract

Purpose: Small-cell lung cancer (SCLC) is difficult to cure. In this study, the SEER database was used to construct a model and explore the potential prognostic factors of SCLC patients.

Methods: The data were sorted out and randomly divided into training cohort and verification cohort. Univariate and multivariate Cox regression were used in the training cohort to analyze the independent prognostic factors, then they be incorporated into the Nomogram model. Using the C-index, calibration algorithm and ROC in conjunction with the risk scores, the model was verified with the verification cohort. Finally, the overall survivals of those factors were evaluated in the total cases.

Results: In the training cohort, we found that age, race, sex, total stage and extension were independent factors which were included in the Nomogram model. C-index(s) that obtained from the training and verification cohorts showed that the model has predictive power. Moreover, the calibration curves and AUC results proved that the model is of great consistency not only in the training cohort but also in the verification cohort. Finally, significant differences in survival were observed among the above-mentioned factors and the overall survivals decreased over time.

Conclusions: Age, race, sex, total stage and extension degree are independent risk factors for overall survival of patients. The Nomogram model can better predict the 1-year, 3-year and 5-year survival probabilities, providing accurate reference for clinical individualized treatment.

Introduction

In 2019, lung cancer had the highest incidence and death rate of all cancers in the United States [1]. Small-cell lung cancer (SCLC) is a malignant pulmonary neuroendocrine tumor originating in bronchial mucosa or glandular epithelium. It is the most malignant and least differentiated lung cancer, accounting for about 13%-20% of lung cancer [2]. Small-cell lung cancer can be divided into oat type, intermediate type, and compound type. Its characteristics mainly include fast growth [3], strong invasive, easy drug resistance, easy recurrence and poor prognosis. Its transfer time is earlier, the scope is wide. Many studies have shown that smoking is a major risk factor for small-cell lung cancer [4].

At present, there are some challenges in the early diagnosis of small-cell lung cancer. It is mainly performed by puncture or biopsy, using Tumor, Node, Metastasis Classification (TNM) system or Veterans Administration (VA) system staging [5]. The clinical treatment of small-cell lung cancer is also difficult. Despite this, patients still have metastasis, drug resistance, and even relapse, resulting in high mortality rates. In order to improve the early diagnosis level of small-cell lung cancer and develop personalized treatment methods, it is quite indispensable to pay attention to the relevant risk factors and predict its survival rate systematically and accurately.

This study systematically analyzed the clinical and pathological data of small-cell lung cancer patients from the Surveillance, Epidemiology, and End Results (SEER) Database (from 1973 to 2015). By

establishing a statistical model that predicts survival in patients, we explored and discussed the clinical application of potential prognostic factors in patients with small-cell lung cancer.

Materials And Methods

Data sources

Patients with small-cell lung cancer diagnosed by pathology in the seer database (1973-2015) were collected by SEER*Stat software. We can extract the patient information including age, race, sex, age of diagnosis, histologic type, tumor site, pathological stage (T, N, and M), surgical site and situation, tumor size, the sequence number of primary tumor, tumor type, lateral, extension, the involvement of lymph nodes, metastasis, death, survival time and state, benign and malignant tumor, and marital status. In order to understand the influencing factors and benefits of surgical treatment, we ignored the possible interference of chemoradiotherapy data. A total of 1207856 patients met the initial requirements.

Data cleaning

Inclusion criteria: (1) The pathological diagnosis was small-cell lung cancer and the Histology/behavior codes are 8041-8042-8043-8044 and 8045 in the International Classification of Diseases for Oncology (ICD-O). (2) The year of diagnosis was 1973-2015. (3) Small-cell lung cancer in the bronchi, bronchioles, or alveoli is the first primary tumor.

Exclusion criteria: (1) Tumor's information is incomplete. (2) Data on patients with unexplained or non-fatal causes of small-cell lung cancer. (3) The primary tumor is not small-cell lung cancer. After screening, a total of 4432 small-cell lung cancer patients were included in the study. We randomly divided these patient data into a training cohort of 1776 (40.07%) and a validation group of 2656 (59.93%) [6-8] with the createDataPartition function in R caret package.

Statistical analysis

After data cleaning, 4432 cases were included in the study. They were randomly divided into a training cohort (1766 cases) and a verification cohort (2656 cases). Univariate Cox regression model was established to analyze the data of the training cohort. The significant variables in univariate analysis were combined with the Cox regression model to construct a multivariate analysis, so as to clarify the independent risk factors affecting the survival rate of small-cell lung cancer in the training cohort. Besides, Nomogram was plotted as a survival prediction model [9, 10]. Then Calibration plots were drawn to evaluate the prediction models for 1-year, 3-year, and 5-year survival rates [10]. By adding meaningful variables: age, gender, race, total staging, and extension into the calculation of each patient's risk score, the risk score of each patient with small-cell lung cancer was calculated and a receiver operating characteristic (ROC) curve assessment model was produced. Meanwhile, the area under the curve (AUC) of ROC of the 1-year, 3-year and 5-year survival rate was calculated to evaluate the accuracy of the model prediction.

In addition, the Cox regression model with multiple factors was constructed again in the validation group. The C index, calibration chart and AUC evaluation were verified.

At last, the risk score with total data (including training cohort and verification cohort) was calculated. The relationship between survival time and survival probability were drawn respectively under the two conditions of high risk and low risk. The overall survivals of meaningful multivariables (age, sex, race, pathological stage, and extension) which came from the above model were also analyzed.

All the above were performed by R language (v3.5.3), and the difference was statistically significant when $P < 0.05$. The whole process can be seen in Fig. S1.

Results

Patient characteristics

As shown in Fig. 1, after data cleaning, 4432 small-cell lung cancer patients were screened, including 1776 in the training cohort and 2656 in the validation group. Clinical and pathological characteristics of patients in the training cohort, validation group and pre-cleaning group are shown in Table 1.

Univariate Cox analysis results of survival influencing factors

Univariate Cox regression analysis was performed on 1776 small-cell lung cancer patients in the training cohort. Clinical pathologic factors of single factor survival analysis results showed that age, race, sex, the degree of differentiation, N, M and total stage, surgery of primary site, tumor size, extension, involvement the lymph nodes, metastasis, death from cancer and non-cancer causes, and sequence of primary tumor, tumor number, and age of diagnosis is associated with survival time and survival ($P < 0.05$). There was no correlation between T stage, operation or no operation, marital status, lateral, survival time and survival status of patients ($P > 0.05$) (Table S1).

Multivariate Cox analysis results of survival influencing factors

The statistically significant variables in the univariate Cox regression analysis were included in the multivariate analysis. The results showed that the patients were aged 65-69, aged 75-79, aged 80-84, aged ≥ 85 , white, male, total stage II, tumor invasion range correlated with survival time and survival status (Table 2).

The Nomogram model building

Based on the results of the multivariate analysis, the Cox regression model was constructed again for the above meaningful variables. They are included in the Nomogram plot and assigned as the point in the Nomogram according to the results of the Cox regression model. The total score can be obtained by adding the single score value of 5 variables. The overall score corresponds to the survival axis values,

which can predict the 1-year, 3-year, and 5-year survival rates of small-cell lung cancer patients. The higher the overall score, the higher the survival rate, and vice versa. Detailed results are shown in Fig. 2.

The Nomogram model was tested by training cohort

C-index was used to evaluate the discrimination between the model and the real value of the training cohort. The results showed that the c-index was 0.6817, indicating that the model was acceptable and capable of prediction.

The results of the model evaluation using the calibration diagram are shown in Fig. 3A~C respectively. Three Numbers were taken on average in the sample number, and the fitting coefficient was set to 100 to calculate the actual 1-year, 3-year and 5-year survival rates corresponding to those predicted by Nomogram. The evaluation results show that the calibration curves of the 1-year, 3-year and 5-year survival prediction of the training cohort are close to the ideal 45° dashed line, indicating that the predicted values are in good consistency with the actual observed values.

ROC curve prediction was used to evaluate the model. The significant variables in the multivariate analysis, such as age, sex, race, total stage, and extension, were included in the calculation of the risk score of each patient. The ROC curve was prepared in combination with the risk score and the AUC was calculated (Fig. 4A~C). The results showed that the 1-year survival rate prediction (AUC=0.733), 3-year survival rate prediction (AUC=0.754) and 5-year survival rate prediction (AUC=0.743) in the training cohort were of moderate accuracy, indicating that the accuracy of survival rate prediction was relatively high. There was good consistency between the predicted value and the actual value.

Verifying the prediction model with the verification cohort

In the verification cohort, the Cox regression model of multiple factors was built first. Then the c-index of the training cohort was verified. The result of c-index is 0.6778, which conforms to the accuracy of model prediction of the training cohort and has the ability of prediction.

It is verified that the calibration diagram of the modeling group is accurate in evaluating the model. The evaluation results of the validation group on the model are shown in Fig. 3D~F. Evaluation results show that the one-year, three-year and five-year survival prediction correction curves of the verification cohort are similar to those of the training cohort, with an ideal 45° dotted line. They indicate that the calibration of the training cohort is relatively accurate.

The ROC curve of the training cohort was verified to evaluate the model. The risk score of each patient in the validation group was calculated. Then the ROC curve and the area under the curve were calculated by combining the risk scores (Fig. 4D~F). The results showed that the 1-year survival rate prediction, 3-year survival rate prediction, and 5-year survival rate prediction of the validation group all belonged to the range of moderate accuracy, indicating that the ROC curve of the training cohort was relatively accurate in evaluating the model.

The overall survival evaluation in the total samples

All patient data from the modeling and validation groups were integrated to obtain the total risk score. The survival curves of high risk and low risk were plotted (Fig. 5A). Based on the Cox regression model established above, the effects of 5 meaningful variables (age, gender, race, total staging and leaching degree) in the multivariate analysis on the survival of patients were analyzed, and the survival curve of each variable was plotted (Fig. 5B~F).

In the survival analysis, the median survival time of small-cell lung cancer was 7 (0-71) months, and the average survival time was (11.26 ± 13.09) months. In the one-year survival rate statistics, the one-year survival rate gradually decreased with the increase of age, from 49.2% to 19.67%. In addition to 50-54 years survival rate (49.7%) was slightly higher than on slightly higher than a group. The 3-year survival rate also declined gradually, from 16.3% to 5.5%, except for the 65-69 years-old survival rate and the 50-54 years-old survival rate, which were slightly higher than the previous age group. Although there is a lack of data on 5-year survival, it is generally declining. The survival rate of black people (41.14%) was higher than that of white people (36.79%) and other people (32.61%). The survival rate of small-cell lung cancer in women (40.53%) was higher than that in men (33.32%). In tumor stage, the survival rate showed a general trend of decreasing with the higher stage, from 76.3% to 22.59%, among which stage IIA (73.0%), stage III (66.7%) and stage IIIB (52.6%) were slightly higher than the previous stage. In terms of the degree of an extension, the survival rate decreased gradually with the increase of the range, from 44.4% to 32.0% (Table S2).

Discuss

This seer study is mainly to analyze the influential factors affecting the survival time and state of patients with small-cell lung cancer in the data of 4432 patients. It finds that age, race, sex, total stage, and extension are risk factors related to survival time and survival status in multivariate analysis. Some studies suggest that the extent of disease, performance status and LDH serum levels are independent prognostic factors for survival, which is inconsistent with the results of this study [11]. Such consequences may be related to the different range and mode of data selection.

(a). In the statistics of 1-year survival rate, 3-year survival rate and 5-year survival rate, generally, the survival rate decreases gradually with the increase of age group. The reasons may be related to physical quality, physical state at the time of treatment, and treatment strategies. In clinical practice, with the increase of age, the more complications patients have, the more conservative treatment strategies doctors may adopt, leading to unsatisfactory treatment effect [12]. Older patients are also likely to be less willing to cooperate with treatment. Most patients aged 50-54 are in the middle of life and in good physical condition, have fewer complications [13] so that they can be able to withstand the side effects brought by chemotherapy and radiotherapy [14]. In conclusion, the patients aged 50-54 have a good early treatment effect and a positive attitude towards treatment.

(b). Although the United States is a multiracial country, it is still predominantly white. White people also have higher average income and education levels than blacks and others, according to the U.S. census bureau [15]. Therefore, whites have a higher awareness of health care and higher affordability of treatment. There are still deficiencies in the racial survival statistics, with seer data lacking in blacks and others. The results would be more convincing if the number of black patients and others were increased.

(c). Survival rates are significantly higher in women than in men, which is consistent with Eskandar et al.'s study of the effects of gender on small-cell lung cancer [16, 17]. The reason may have something to do with men's smoking. Moreover, women are less likely to smoke in general, to a lesser extent than men. As a result, small-cell lung cancer is less severe in women than in men.

(d). As the total stage increased, the 1-year, 3-year, and 5-year survival analyses showed a decrease in volatility. It may be related to the treatment strategies adopted by patients. In general, the smaller the stage, the lighter the disease, the better the physical state, the better the effect of chemotherapy, radiotherapy and other treatment means. However, patients with high stage are weak and the cancer is likely to have metastasized, so it is difficult to take the treatment with large side effects and the curative effect is not satisfactory.

(e). As the degree of tumor invasion increases, the scope and extent of tumor damage to surrounding tissues will increase. Likewise, the possibility of distant tumor metastasis will also increase. Even when patients adopt aggressive treatment strategies, the results are less and less effective, so the survival rate is inevitably lower.

(f). The degree of differentiation, N, M and total stage, surgery of primary site [18], tumor size, involvement the lymph nodes [5, 19], metastasis [20], death from cancer and non-cancer causes, and sequence of the primary tumor, tumor number, and age of diagnosis is meaningful in univariate analysis, but meaningless in the multifactor analysis. The first possibility is that these factors are not strongly correlated with the survival time and status of patients [21]. The second possibility is that there are cross-influences among various factors, leading to the weakened correlation between factors and results. The third possibility is that these factors are affected by some confounding factors, leading to errors in the statistical data.

In this study, on the whole, the SEER database adopted in this study not only has a large amount of data and rich contents, but also has a large time span of collected patient data. And the patients were from all over the United States, reducing the statistical bias of patient data from a single institution. These advantages make the results of this study have a large sample size, so the results are more convincing. However, the seer database still has some drawbacks. Firstly, the data records of many aspects of patients in the database are unknown, which greatly reduces the sample size that can be finally included in the analysis after data cleaning. Secondly, the patients in the database are all American patients, white patients are the majority, and the sample size of black and other races is insufficient. Hence, the survival prediction model constructed is not suitable for Asia or the world. Third, there were no relevant records of the patients' physical health conditions before treatment, such as basic diseases, smoking history and

tumor marker examination results [22-24]. No details were given about the treatment methods [25, 26], the drugs used to treat it and radiotherapy timing [27]. Adjuvant therapy and quality of life after treatment were also not mentioned. In the following in-depth studies, more specific details of the factors related to the survival of patients should be explored, such as specific treatment strategies based on different age or physical conditions. At the same time, we will refer to more different databases to study the specific impact of different regions and different races on survival.

Conclusion

On the whole, by identifying and analyzing the independent influencing factors of small-cell lung cancer, this study obtained the effect of different influencing factors on the survival time and survival probability of patients. Besides, a Nomogram was drawn to make a more accurate prediction of 1-year, 3-year and 5-year survival rates for different patients. Patients can also make the most appropriate treatment decisions based on prognosis, reducing the incidence of incomplete treatment and inappropriate treatment timing.

The final results are of great significance for objective analysis of patients' conditions, formulation of personalized clinical treatment strategies, and even prediction of patients' prognosis. However, there are still some limitations in this study. The influencing factors of small cell lung cancer explored and obtained in the research process are all known, and it has not been analyzed from a new perspective or drawn new conclusions in biology, which is lack of innovation.

In this study, Cox regression model, Nomogram graph and ROC curve model were constructed based on seer database, which inevitably had undeniable defects. The validation of each model also comes from internal data. Verification of data from other sources will certainly increase its credibility. We should continue to collect data on related diseases in other countries and regions, so that we could perfect research results and data analysis, as well as reduce objective errors and imperfections.

Declarations

Competing Interests

The authors declare that they have no competing interests.

Consent for publication

We agree to publish this paper in BMC.

Authors' contributions

ST performed the statistical analyses and wrote the first draft of the manuscript. XZ checked the statistical accuracy as an expert in statistics. YK performed the literature search and discussed the results. YK, SL and XZ revised and edited the final version of the manuscript. All authors read and approved the final manuscript.

Ethical approval

The work was approved by Guangdong Medical University Ethics committee (YS2019131), and in accordance with the Declaration of Helsinki of the World Medical Association. Informed consent forms are not required for patient data extracted from SEER database.

Data Availability

The raw data used and/or analyzed during the current study are available from the first author (ST, 16101200587@gdmu.edu.cn) on reasonable request.

Funding

This work was supported partly by National Natural Science Foundation of China (81541153 and J18111211); Guangdong Science and Technology Department (2016A050503046 and 2016B030309002); The Public Service Platform of South China Sea for R&D Marine Biomedicine Resources (GDMUK201808); Southern Marine Science and Engineering Guangdong Laboratory Zhanjiang (ZJW-2019-07). The funders had no role in the design of the study; the collection, analysis, and interpretation of the data; the writing of the manuscript; and the decision to submit the manuscript for publication.

References

1. Siegel RL, Miller KD, Jemal A: **Cancer statistics, 2020**. *CA Cancer J Clin* 2020, **70**(1):7-30.
2. Salcido CD, Larochele A, Taylor BJ, Dunbar CE, Varticovski L: **Molecular characterisation of side population cells with cancer stem cell-like characteristics in small-cell lung cancer**. *Br J Cancer* 2010, **102**(11):1636-1644.
3. Vostrejs M, Moran PL, Seligman PA: **Transferrin synthesis by small cell lung cancer cells acts as an autocrine regulator of cellular proliferation**. *J Clin Invest* 1988, **82**(1):331-339.
4. Hamilton G, Rath B: **Smoking, inflammation and small cell lung cancer: recent developments**. *Wien Med Wochenschr* 2015, **165**(19-20):379-386.
5. Zhu X, Luo H, Xu Y: **Transcriptome analysis reveals an important candidate gene involved in both nodal metastasis and prognosis in lung adenocarcinoma**. *Cell Biosci* 2019, **9**:92.
6. Wang CY, Yang J, Zi H, Zheng ZL, Li BH, Wang Y, Ge Z, Jian GX, Lyu J, Li XD *et al*: **Nomogram for predicting the survival of gastric adenocarcinoma patients who receive surgery and chemotherapy**. *BMC Cancer* 2020, **20**(1):10.
7. Hu Y, You S, Yang Z, Cheng S: **Nomogram predicting survival of hepatocellular carcinoma with portal vein tumour thrombus after curative resection**. *ANZ J Surg* 2019, **89**(1-2):E20-E25.
8. Wang C, Yang C, Wang W, Xia B, Li K, Sun F, Hou Y: **A Prognostic Nomogram for Cervical Cancer after Surgery from SEER Database**. *J Cancer* 2018, **9**(21):3923-3928.

9. Etzel CJ, Kachroo S, Liu M, D'Amelio A, Dong Q, Cote ML, Wenzlaff AS, Hong WK, Greisinger AJ, Schwartz AG *et al*: **Development and validation of a lung cancer risk prediction model for African-Americans.** *Cancer Prev Res (Phila)* 2008, **1**(4):255-265.
10. Yap WK, Shih MC, Kuo C, Pai PC, Chou WC, Chang KP, Tsai MH, Tsang NM: **Development and Validation of a Nomogram for Assessing Survival in Patients With Metastatic Lung Cancer Referred for Radiotherapy for Bone Metastases.** *JAMA Netw Open* 2018, **1**(6):e183242.
11. Hermes A, Waschki B, Gatzemeier U, Reck M: **Characteristics, treatment patterns and outcomes of patients with small cell lung cancer—a retrospective single institution analysis.** *Lung Cancer* 2011, **71**(3):363-366.
12. Ludbrook JJ, Truong PT, MacNeil MV, Lesperance M, Webber A, Joe H, Martins H, Lim J: **Do age and comorbidity impact treatment allocation and outcomes in limited stage small-cell lung cancer? a community-based population analysis.** *Int J Radiat Oncol Biol Phys* 2003, **55**(5):1321-1330.
13. Sas-Korczynska B, Luczynska E, Kamzol W, Sokolowski A: **Analysis of risk factors for pulmonary complications in patients with limited-stage small cell lung cancer : A single-centre retrospective study.** *Strahlenther Onkol* 2017, **193**(2):141-149.
14. Gregor A, Drings P, Rinaldi M, Schuster L, Burghouts J, Postmus PE, Dalesio O, Kirkpatrick A, Hocht Boes G, Van Zandwijk N: **Acute toxicity of alternating schedule of chemotherapy and irradiation in limited small-cell lung cancer in a pilot study (08877) of the EORTC Lung Cancer Cooperative Group.** *Ann Oncol* 1995, **6**(4):403-405.
15. Akee R, Jones MR, Porter SR: **Race Matters: Income Shares, Income Inequality, and Income Mobility for All U.S. Races.** *Demography* 2019, **56**(3):999-1021.
16. Eskandar A, Ahmed A, Daughtey M, Kenderian S, Mahdi F, Khan A: **Racial and sex differences in presentation and outcomes of small cell lung cancer in the United States: 1973 to 2010.** *Chest* 2015, **147**(4):e164-e165.
17. Lim JH, Ryu JS, Kim JH, Kim HJ, Lee D: **Gender as an independent prognostic factor in small-cell lung cancer: Inha Lung Cancer Cohort study using propensity score matching.** *PLoS One* 2018, **13**(12):e0208492.
18. Anraku M, Waddell TK: **Surgery for small-cell lung cancer.** *Semin Thorac Cardiovasc Surg* 2006, **18**(3):211-216.
19. Urban T, Chastang C, Vaylet F, Mathieu M, Leclerc P, Paillot D, Lebeau B: **Prognostic significance of supraclavicular lymph nodes in small cell lung cancer: a study from four consecutive clinical trials, including 1,370 patients. "Petites Cellules" Group.** *Chest* 1998, **114**(6):1538-1541.
20. Wang H, Han X, Guo J, Wang Z: **Characteristics and survival difference of clinical tumor size 0 extensive-stage small cell lung cancer with different metastasis pattern.** *J Thorac Dis* 2018, **10**(9):5414-5420.
21. Zou Q, Chen L, Huang T, Zhang Z, Xu Y: **Machine learning and graph analytics in computational biomedicine.** *Artif Intell Med* 2017, **83**:1.

22. Huang CH, Wick JA, Sittampalam GS, Nirmalanandhan VS, Ganti AK, Neupane PC, Williamson SK, Godwin AK, Schmitt S, Smart NJ *et al*: **A multicenter pilot study examining the role of circulating tumor cells as a blood-based tumor marker in patients with extensive small-cell lung cancer.** *Front Oncol* 2014, **4**:271.
23. Brenner S, Klameth L, Riha J, Scholm M, Hamilton G, Bajna E, Ausch C, Reiner A, Jager W, Thalhammer T *et al*: **Specific expression of OATPs in primary small cell lung cancer (SCLC) cells as novel biomarkers for diagnosis and therapy.** *Cancer Lett* 2015, **356**(2 Pt B):517-524.
24. Zhu X, Lin MCM, Fan W, Tian L, Wang J, Ng SS, Wang M, Kung H, Li D: **An intronic polymorphism in GRP78 improves chemotherapeutic prediction in non-small cell lung cancer.** *Chest* 2012, **141**(6):1466-1472.
25. Wakeam E, Acuna SA, Leighl NB, Giuliani ME, Finlayson SRG, Varghese TK, Darling GE: **Surgery Versus Chemotherapy and Radiotherapy For Early and Locally Advanced Small Cell Lung Cancer: A Propensity-Matched Analysis of Survival.** *Lung Cancer* 2017, **109**:78-88.
26. Luchtenborg M, Riaz SP, Lim E, Page R, Baldwin DR, Jakobsen E, Vedsted P, Lind M, Peake MD, Mellempgaard A *et al*: **Survival of patients with small cell lung cancer undergoing lung resection in England, 1998-2009.** *Thorax* 2014, **69**(3):269-273.
27. Scotti V, Meattini I, Franzese C, Saieva C, Bertocci S, Meacci F, Furfaro I, Scartoni D, Cecchini S, Desideri I *et al*: **Radiotherapy timing in the treatment of limited-stage small cell lung cancer: the impact of thoracic and brain irradiation on survival.** *Tumori* 2014, **100**(3):289-295.

Tables

Table 1. Demographic characteristics of SCLC patients from SEER (1973-2015)

Variants	Before data cleaning (n=1207856, %)	After data cleaning(4452)	
		Training cohort (n1=1776, %)	Validation cohort (n2=2656, %)
Age(years)			
0-49	63161(5.23%)	68(3.83%)	103(3.88%)
50-54	72771(6.02%)	137(7.71%)	189(7.12%)
55-59	115937(9.60%)	204(11.49%)	312(11.75%)
60-64	159287(13.19%)	264(14.86%)	426(16.04%)
65-69	197350(16.34%)	357(20.10%)	520(19.58%)
70-74	204589(16.94%)	302(17.00%)	450(16.94%)
75-79	182102(15.08%)	236(13.29%)	370(13.93%)
80-84	126825(10.50%)	134(7.55%)	203(7.64%)
85+	85834(7.11%)	74(4.17%)	83(3.12%)
Race			
Black	128503(10.64%)	133(7.49%)	269(10.13%)
White	1011292(83.73)	1576(88.74%)	2304(86.75%)
Other	66556(5.51%)	67(3.77%)	83(3.12%)
Unknown	1505(0.12%)		
Sex			
Female	526891(43.62%)	889(50.06%)	1387(52.22%)
Male	680965(56.38%)	887(49.94%)	1269(47.78%)
Grade			
Grade I(Well differentiated)	49189(4.07%)	12(0.68%)	17(0.64%)
Grade II(Moderately differentiated)	155205(12.85%)	20(1.13%)	30(1.13%)
Grade III(Poorly differentiated)	287860(23.83%)	698(39.30%)	990(37.27%)
Grade IV(Undifferentiated; anaplastic)	98042(8.12%)	1046(58.90%)	1619(60.96%)
Unknown	617560(51.13%)		
Stage			
IA	40610(3.36%)	76(4.28%)	139(5.23%)
IB	19953(1.65%)	40(2.25%)	81(3.05%)
II	204(0.01%)	1(0.06%)	
	10948(0.91%)	67(3.77%)	85(3.20%)

IIA	11561(0.96%)	39 (2.20%)	52(1.96%)
IIB	3570(0.30%)	2(0.11%)	5(0.19%)
III	35307(2.92%)	309(17.40%)	432(16.27%)
IIIA	16723(1.38%)	182(10.25%)	288(10.84%)
IIIB	144424(11.96%)	1060(59.68%)	1574 (59.26%)
IV	924556(76.55%)		
Unknown			
Stage_T	1810(0.15%)	13(0.73%)	19(0.72%)
T0	37157(3.08%)	125(7.04%)	210(7.91%)
T1a	26211(2.17%)	116(6.53%)	189(7.12%)
T1b	336(0.03%)		3(0.11%)
T1NOS	50579(4.19%)	343(19.31%)	463(17.43%)
T2a	18309(1.52%)	142(8.00%)	240(9.04%)
T2b	8372(0.69%)	8(0.45%)	7(0.26%)
T2NOS	55290(4.58%)	430(24.21%)	583(21.95%)
T3	61545(5.10%)	599(33.73%)	942(35.47%)
T4	948247(78.51%)		
Unknown			
Stage_N	118858(9.84%)	364(20.50%)	534(20.11%)
N0	23367(1.93%)	166(9.35%)	255(9.60%)
N1	98312(8.14%)	917(51.63%)	1383(52.07%)
N2	35548(2.94%)	329(18.52%)	484(18.22%)
N3	931771(77.14%)		
Unknown			
Stage_M	163491(13.54%)	716(40.32%)	1082(40.74%)
M0	39995(3.31%)	184(10.36%)	285(10.73%)
M1a	101911(8.44%)	858(48.31%)	1268(47.74%)
M1b	2518(0.21%)	18(1.01%)	21(0.79%)
M1NOS	899941(74.51%)		
Unknown			
Size(mm)	192171(15.91%)	488(27.48%)	778(29.29%)

<=30	127739(10.58%)	482(27.14%)	680(25.60%)
>30,<=50	118472(9.81%)	690(38.85%)	991(37.31%)
>50,<=100	13661(1.13%)	116(6.53%)	207(7.79%)
>100	755813(62.57%)		
Unknown			
Extension	190020(15.73%)	575(32.38%)	792(29.82%)
<=100	71004(5.88%)	231(13.01%)	402(15.14%)
>100,<=300	82565(6.84%)	355(19.99%)	501(18.86%)
>300,<=500	271275(22.46%)	615(34.63%)	961(36.18%)
>500,<=999	592992(49.09%)		
Unknown			
Involvement of lymph nodes	227488(18.83%)	364(20.50%)	534(20.11%)
None	387376(32.07%)	1412(79.50%)	2122(79.89%)
Lymphatic involvement	592992(49.09%)		
Unknown			
Metastasis	291112(24.10%)	702(39.53%)	1062(39.98%)
No metastasis	323752(26.80%)	1074(60.47%)	1594(60.02%)
Distant metastasis	592992(49.09%)		
Unknown			
Cancer death	223996(18.54%)	484(27.25%)	770(28.99%)
Alive or dead of other cause	732669(60.66%)	1292(72.75%)	1886(71.01%)
Dead (attributable to this cancer)	251191(20.80%)		
Unknown			
Other Death	838213(69.40%)	1679(94.54%)	2512(94.58%)
Alive or dead due to cancer	118452(9.81%)	97(5.46%)	144(5.42%)
Dead (attributable to causes other than this cancer)	251191(20.80%)		
Unknown			
Status	142669(11.81%)	387(21.79%)	626(23.57%)
Alive	1065187(88.19%)	1389(78.21%)	2030(76.43%)
Dead			

Sequence number of primary tumors	906913(75.08%)	1724(97.07%)	2564(96.54%)
One primary only	58351(4.83%)	52(2.93%)	92(3.46%)
1st of 2 or more primaries	242535(20.08%)		
Others	57(0.01%)		
Unknown			
Total number of in situ/malignant tumors for patient	910986(75.42%)	1729(97.35%)	2576(96.99%)
1	231284(19.15%)	46(2.59%)	74(2.79%)
2	51746(4.28%)	1(0.06%)	6(0.23%)
3	13783(1.14%)		
Others	57(0.01%)		
Unknown			
Age of diagnosis	63161(5.23%)	68(3.83%)	103(3.88%)
0-49	386132(31.97%)	677(38.12%)	1046(39.38%)
50-65	758563(62.80%)	1031(58.05%)	1507(56.74%)
>=66	130487(10.80%)	256(14.41%)	353(13.29%)
Marital status	639310(52.93%)	927(52.20%)	1366(51.43%)
Single (never married)	390924(32.37%)	593(33.39%)	937(35.28%)
Married or partner	47135(3.90%)		
Separated, divorced or widowed			
Unknown	1108346(91.76%)	1755(98.82%)	2622(98.72%)
Lateral	17877(1.48%)	6(0.34%)	10(0.38%)
Left, right, or unilateral	81633(6.76%)	15(0.84%)	24(0.90%)
Bilateral			
paired site or not a paired site	649077(53.74%)	1636(92.12%)	2426(91.34%)
Surgery of Primary Site	179763(14.88%)	140(7.88%)	230(8.66%)
No surgery of primary site	379016(31.38%)		
Surgery on primary site			
Blank(s), unknown or NOS			
Whether or not the patient has	629128(52.09%)	1743(98.14%)	2598(97.82%)

surgery	129389(10.71%)	33(1.86%)	58(2.18%)
No surgical	449339(37.20%)		
Surgical			
Blank(s), unknown			

Table 2. Multivariate Cox regression analysis results of SCLC
 inf=infinite, NA=no answer

Figures

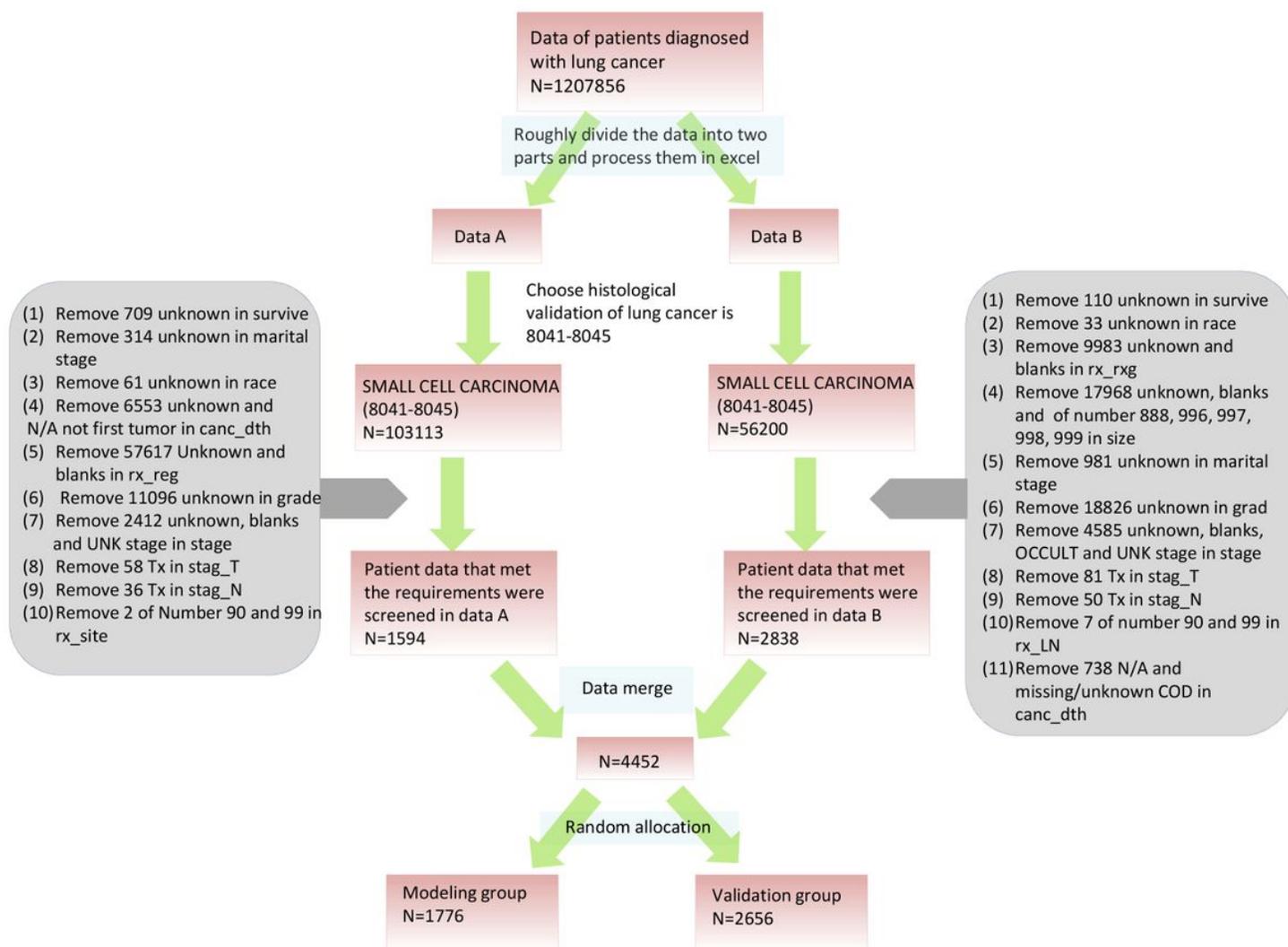


Figure 1

Variables	HR	95%CI	P	SE	Small-cell lung cancer data cleaning process. Because the amount of small-cell lung cancer data is too large, exceeding the processing power of a single excel. Therefore the whole data were split into two groups for cleanup. We can clear the unknown, blank, and unclassified data in each cohort. The data that were not relevant to the study or of no research significance were then purged. Finally, the two groups of processed data are merged. The different types of data are numbered in line with the needs of the research content, which is helpful for the following data analysis.
Age					
Age0-49	1	Reference			
Age50-54	1.02	0.65-1.61	0.926	0.093	
Age55-59	1.10	0.71-1.70	0.682	0.410	
Age60-64	1.23	0.80-1.89	0.351	0.932	
Age65-69	1.38	1.02-1.87	0.037	2.086	
Age70-74	1.34	0.99-1.81	0.057	1.903	
Age75-79	1.53	1.13-2.09	0.007	2.711	
Age80-84	1.64	1.17-2.29	0.004	2.900	
Age>=85	2.38	1.64-3.46	5.680x10 ⁻⁶	4.538	
Race					
Black	1	Reference		2.390	
White	1.29	1.05-1.60	0.017	1.583	
Others	1.32	0.94-1.85	0.113		
Sex					
Female	1	Reference		3.026	
Male	1.19	1.06-1.32	0.002		
Grade					
Grade I	1	Reference		-0.111	
Grade II	9.43x10 ⁻¹	0.34-2.65	0.912	0.726	
Grade III	1.35	0.60-3.06	0.468	0.759	
Grade IV	1.37	0.61-3.09	0.448		
Stage					
IA	1	Reference		1.566	
IB	1.53	0.90-2.61	0.117	2.089	
II	8.72	1.14-66.55	0.037	0.691	
IIA	1.19	0.72-1.97	0.490	1.490	
IIB	1.54	0.87-2.70	0.136	0.000	
III	2.47	0.00-Inf	1.000	1.720	
IIIA	1.43	0.95-2.15	0.086	1.750	
IIIB	1.49	0.95-2.32	0.080	1.477	
IV	1.90	0.81-4.43	0.140		
Stage N					
N0	1	Reference		-1.046	
N1	8.77x10 ⁻¹	0.69-1.12	0.296	0.774	
N2	1.07	0.90-1.28	0.439	0.457	
N3	1.05	0.85-1.29	0.648		
Stage M					
M0	1	Reference		-1.675	
M1a	6.44 x10 ⁻¹	0.38-1.08	0.094	0.123	
M1b	1.03	0.63-1.69	0.902	NA	
M1NOS	NA	NA	NA		
Surgery of Primary Site					
No surgery of primary site	1	Reference		-1.246	
Surgery on primary site	0.84	0.64-1.11	0.213		
Size					
<=30	1	Reference		-0.073	
>30,<=50	0.99	0.85-1.16	0.942	-1.009	
>50,<=100	0.92	0.79-1.08	0.313	-0.595	
>100	0.93	0.73-1.19	0.552		
Extension					
<=100	1	Reference		0.840	
>100,<=300	1.080	0.90-1.29	0.401	0.331	
>300,<=500	1.028	0.87-1.21	0.741	3.078	
>500,<=999	1.266	1.09-1.47	0.002		
Involvement of lymph Nodes					
None	1	Reference		NA	
Lymphatic involvement	NA	NA	NA		
Metastasis					
No metastasis	1	Reference		1.703	
Distant metastasis	1.67	0.93-3.03	0.088		
Cancer death					
Alive or dead of other cause					

Dead (attributable to this cancer)		Reference		
Other Death	1			0.027
Alive or dead due to cancer		0.00-inf	0.978	
Dead (attributable to causes other than this cancer)	3.87x10 ⁸			
Sequence number of primary tumors	1	Reference		0.028
One primary only		0.00-inf	0.978	
1st of 2 or more Primaries	4.05x10 ⁸			
Total number of in situ/ malignant tumors for patient		Reference		-0.515
1	1	0.17-2.84	0.607	
2	0.69			
3				
Age of diagnosis				
0-49		Reference		0.042
50-65		0.24-4.38	0.967	0.000
>=66	1	0.00-Inf	1.000	
	1.03			
	0.46	Reference		-0.589
	1	0.67-1.24	0.556	NA
	0.91	NA	NA	
	NA			

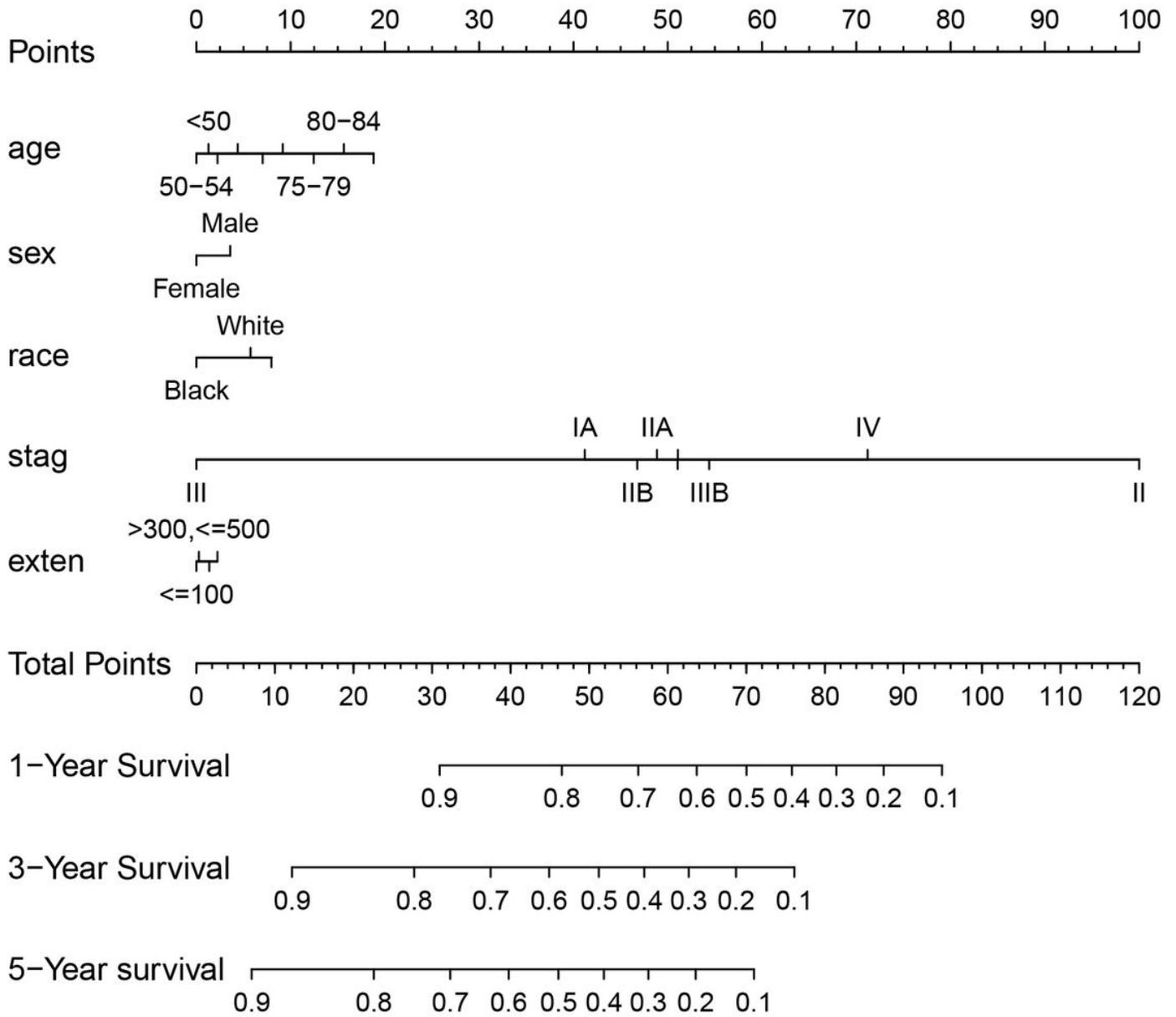


Figure 2

The nomogram model of small-cell lung cancer patients. The cleaned SCLC data were randomly divided into the modeling group and the validation group. First, univariate and multivariate regression analyses were used to analyze the data of the modeling group. Then the meaningful variables filtered out were to construct a Cox regression model. The degree of prediction can be evaluated with c-index, Calibration and ROC curve. The model was applied to the validation group to verify the degree to which the model predicted its data. We can record the risk scores of the modeling group and the validation group respectively. Finally, the model was applied to all small-cell lung cancer data to calculate the total risk score and study the impact of meaningful variables on survival.

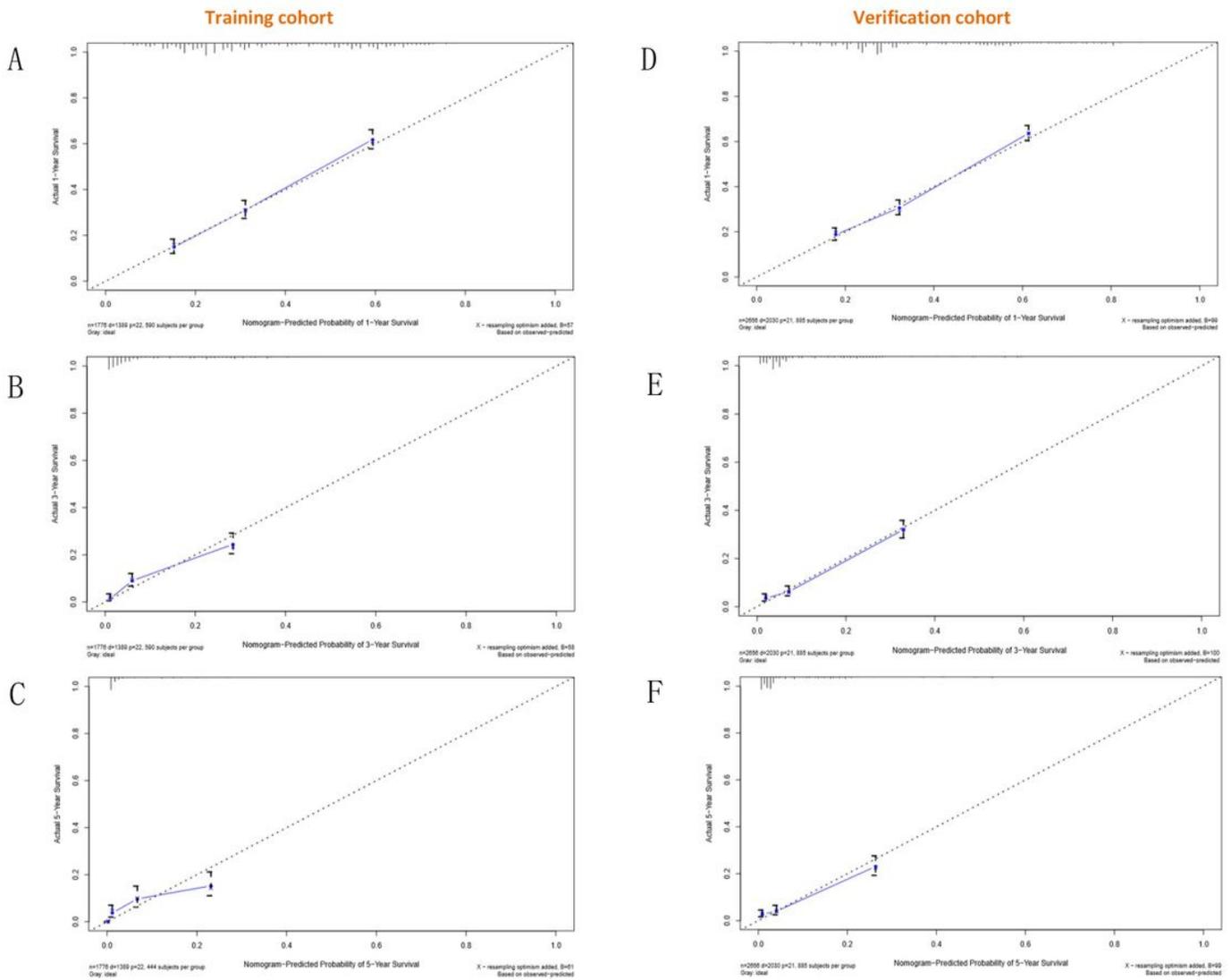
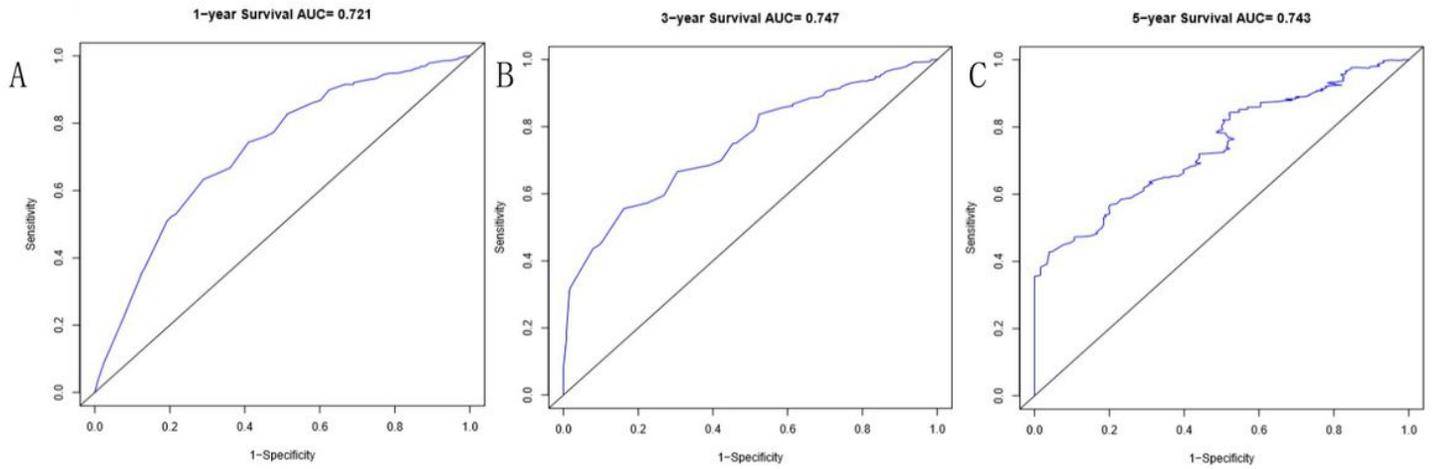


Figure 3

Calibration curves of small-cell lung cancer patients in the training cohort. A. calibration curve predicting 1-year survival (training cohort); B. calibration curve predicting 3-year survival (training cohort); C. calibration curve predicting 5-year survival (training cohort); D. calibration curve predicting 1-year survival (validation cohort); E. calibration curve predicting 3-year survival (validation cohort); F. calibration curve predicting 5-year survival (validation cohort).

Training cohort



Verification cohort

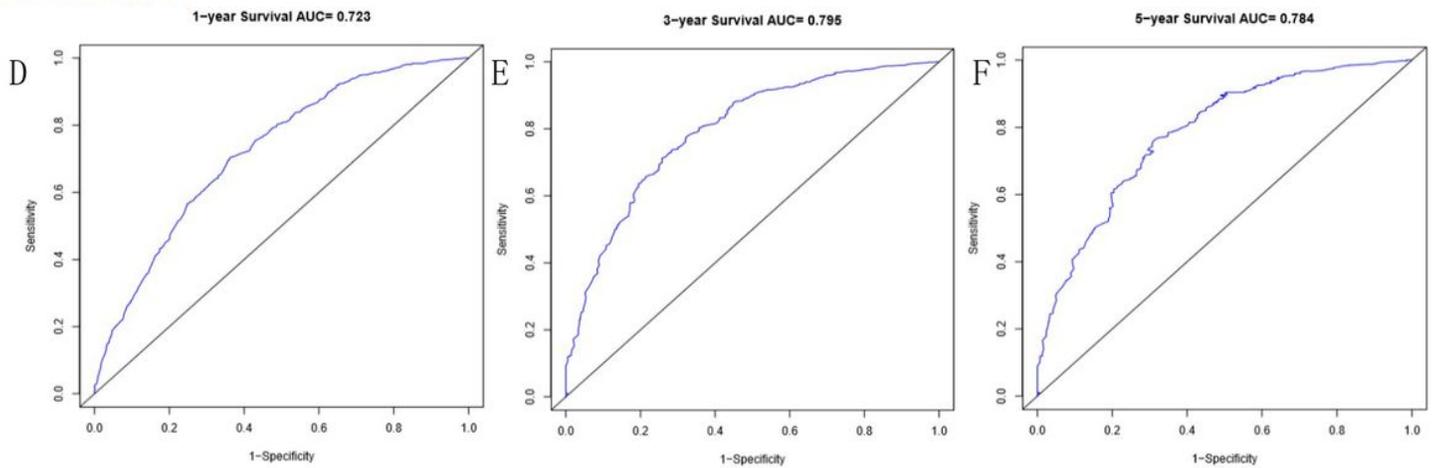


Figure 4

The AUC curve of ROC of SCLC patients. A. 1-year Survival AUC (training cohort); B. 3-year Survival AUC (training cohort); C. 5-year Survival AUC (training cohort); D. 1-year Survival AUC (validation cohort); E. 3-year Survival AUC (validation cohort); F. 5-year Survival AUC (validation cohort).

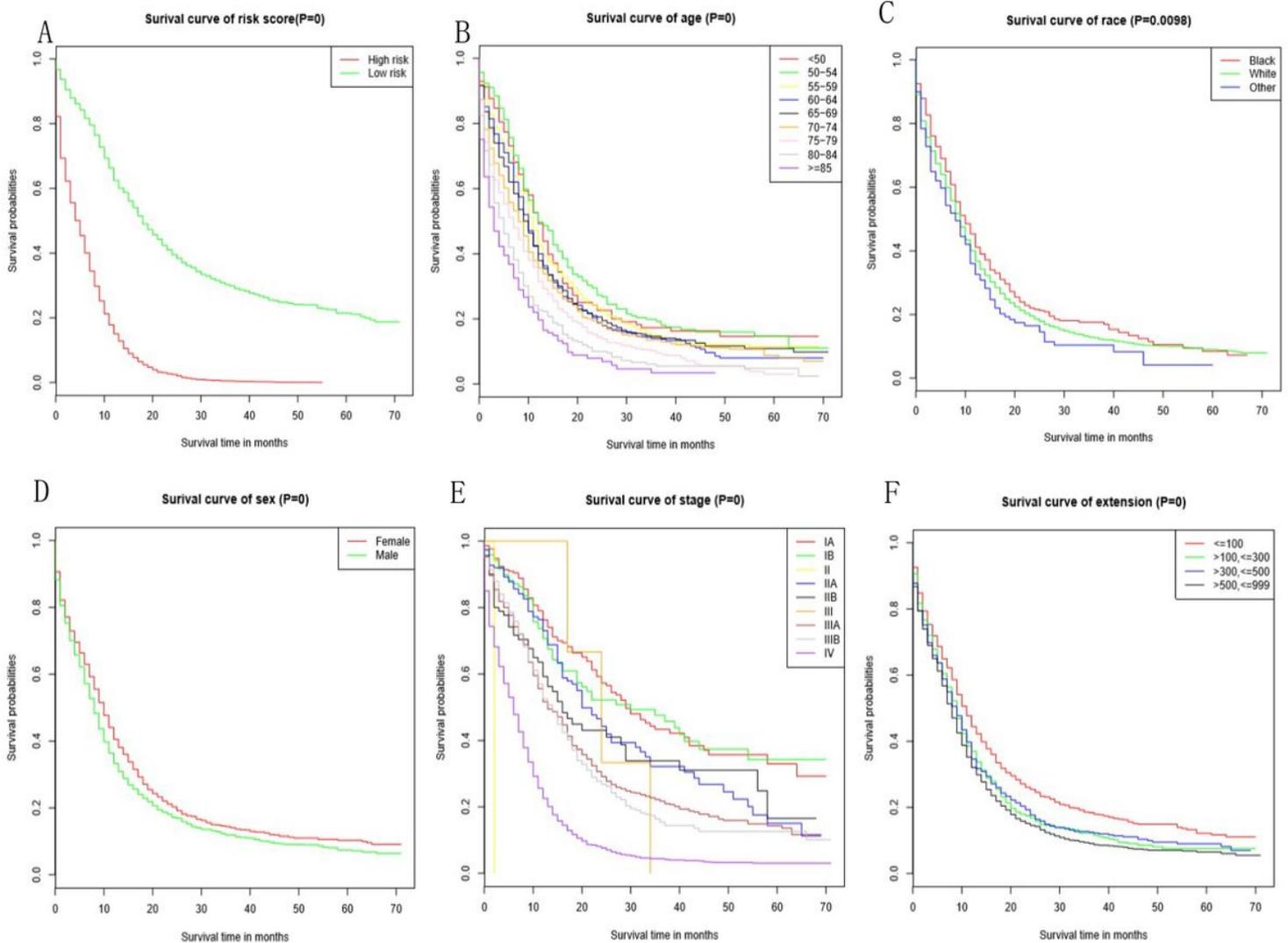


Figure 5

Survival curves of risk score or independent factors in total SCLC patients. (p=0 means P value < 0.001)

A. risk score. The survival probability and survival time of high-risk group were significantly lower than that of low-risk group; B. age. As the growth of the age, the survival time and the chances of survival go down. Nevertheless, the survival curve of patients aged 50-54 was better than that of other age groups; C. race. According to the survival curve, the Blacks have the better survival in SCLC; D. sex. Over time, the possibilities of survival decline. Under the comparison, the possibilities of survival are better for women than for men. E. stage. The lower the staging, the higher the survival time and the survival rate. F. extension. In general, the greater the degree of extension, the lower the patients' survival time and survival probabilities.

Supplementary Files

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

- [supplement11.docx](#)
- [supplement12.docx](#)
- [supplement13.pdf](#)