—Original—

Nomograms to Predict Survival in Patients with Lung Squamous Cell Cancer: A Population-Based Study

Rongjiong Zheng[1], Xiaolong Gu[1], Mingming Wang[1], Meiling Hu[2] and Haiqi Xu[1]

[1]Department of Pulmonology, Ningbo Yinzhou Second Hospital, Ningbo, China
[2]Cixi People's Hospital of Zhejiang Province, Zhejiang, China

Background: This study aimed to identify risk factors affecting cancer-specific survival (CSS) and overall survival (OS) in patients with lung squamous cell carcinoma (LSCC) and to develop nomograms for prognostic prediction in these patients.

Methods: Patients who received an LSCC diagnosis between 2007 and 2013 were selected from the Surveillance, Epidemiology, and End Results (SEER) database. The prognostic effect of each variable on survival was evaluated with Cox regression and Kaplan-Meier analysis, and nomograms were developed to predict 3-, 5-, and 7-year CSS and OS rates.

Results: Data from 23,004 patients with LSCC were analyzed. Nomograms were first developed by using variables that were significantly associated with CSS and OS and then validated by using an internal bootstrap resampling approach, which showed that they had a sufficient level of discrimination, according to the C-index.

Conclusions: The nomograms satisfactorily predicted 3-, 5-, and 7-year CSS and OS rates for patients with LSCC. (J Nippon Med Sch 2019; 86: 336–344)

Key words: lung squamous cell carcinoma, risk factors, nomogram

Introduction

In recent decades, lung cancer, including squamous cell carcinoma, adenocarcinoma, undifferentiated/large cell carcinoma, and small cell lung cancer[1–3], has received close attention and become an important public health issue worldwide[4,5]. Lung cancer, the most common cancer in the world, affected 1.8 million patients and caused 1.6 million deaths in 2012[6]. Lung squamous cell carcinoma (LSCC) represents 30% to 50% of all lung cancers. It is one of the most frequent histological subtypes and is closely related to tumor mortality[7]. Accumulating evidence indicates that LSCC is associated with many chronic diseases and disorders, including diabetes, hypertension, obesity, and metabolic syndrome[8–10], thus dramatically increasing its clinical and economic burden[11]. Because of the seriousness of LSCC, a technically feasible, readily available grading system is needed in order to stratify prognosis.

Nomograms have been used extensively to predict prognoses[12,13]. By integrating important prognostic variables, a nomogram can accurately estimate survival of individual patients. However, few researchers have used crowd data to develop a visual tool for LSCC prognostication. Therefore, we used data from the Surveillance, Epidemiology, and End Results (SEER) database to identify risk factors for cancer-specific survival (CSS) and overall survival (OS) and develop a nomogram to evaluate LSCC prognosis.

Materials and Methods

Data for the present patients with LSCC were retrieved from the SEER database registry of the National Cancer Institute by using SEER*Stat software Version 8.3.5[14]. The SEER database is an aggregation of population-based variables and information on primary tumor characteristics, cancer prevalence, incidence, mortality, and treat-

Correspondence to Meiling Hu, MD, Cixi People's Hospital of Zhejiang Province, Zhejiang, China
E-mail: humeiling826@163.com
https://doi.org/10.1272/jnms.JNMS.2020_86-610
Journal Website (https://www.nms.ac.jp/sh/jnms/)

Table 1 Baseline characteristics of patients

Variables	All patients (N = 23,004)	
	No.	%
Sex		
Female	8,580	37.3
Male	14,424	62.7
Marital status		
Single and Unmarried	3,019	13.1
Married	12,299	53.5
Divorced and Separated	3,277	14.3
Widowed	4,409	19.2
Age, years		
<60	8,580	37.3
≥60	14,424	62.7
Race		
White	19,220	83.6
Black	2,597	11.3
Other[a]	1,187	5.2
Grade		
I	676	2.9
II	9,772	42.5
III	12,310	53.5
IV	246	1.1
TNM stage		
I	7,753	33.7
II	2,257	9.8
III	7,090	30.8
IV	5,904	25.7
Surgery		
No	13,745	59.8
Yes	9,259	40.3
Radiotherapy		
No	20,646	89.7
Yes	2,358	10.3
Chemotherapy		
No	13,071	56.8
Yes	9,933	43.2

[a] Other includes American Indian/Alaska Native, Asian/Pacific Islander, and unknown.

ment, excluding chemotherapy, from 17 geographical regions in the United States[15].

Data from patients with LSCC during the period from January 2007 through December 2013 were analyzed. AJCC Eighth Edition TNM staging was determined on the basis of tumor extension, tumor size, and the criteria for Sixth or Seventh Edition TNM staging[16]. Patients were excluded (1) if information was missing or deficient on survival, histological grade, duration of follow-up, or cause of death, (2) if the diagnosis was based only on autopsy findings or death certificate, (3) if a previous LSCC or more than one primary carcinoma was noted in the patient record, or (4) if age at diagnosis was less than 18 years.

In this study, the following information was obtained for each patient: marital status, race, sex, age, year of diagnosis, TNM stage, histological grade, surgery, chemotherapy, radiotherapy, cause of death, and follow-up information. Then, patients were classified into two groups based on diagnostic age, specifically those younger than 60 years and those older than 60 years. Patients were classified by race as white, black, and other, which included American Indian/Alaska Native, Asian/Pacific Islander, and unknown. The marital status of patients was categorized as single and unmarried, married, divorced and separated, and widowed. TNM staging was categorized on the basis of the system outlined in the AJCC Eighth Edition. The primary endpoints were CSS (interval from an LSCC diagnosis to its censor) and OS (interval from an LSCC diagnosis to death or last follow-up), with no restrictions on cause of death.

Categorical data were displayed in relation to frequency and scale. Univariate and multivariate analyses were performed by using the Cox proportional hazards regression model. All statistical analyses were carried out by using SPSS (Version 17.0, SPSS software, Chicago, IL, USA) and R Version 3.5.1 (http://www.r-project.org/)[17]. In addition, the R packages rms and mstate were used to frame the model and nomogram. The concordance index (C-index), which measures differences in predictive power between observed and predicted results, was used to evaluate the discrimination of the nomograms[18]. A higher C-index value indicates a better ability to discriminate between patients with different survival outcomes[19]. Then, the C-index and calibration curve were obtained by using regression analysis. A calibration plot along a 45-degree line indicates a perfectly calibrated model, in which the predicted probabilities are the same as the actual results. A P value of <0.05 was considered to indicate statistical significance.

The study was performed in accordance with the guidelines of the Declaration of Helsinki. The study protocol and consent form were approved by the Ethics Committee of the Institutional Review Board of Ningbo Yinzhou Second Hospital, Ningbo 315192, China.

Results

Patient Demographic and Pathological Characteristics

Data from 23,004 eligible patients with LSCC diagnoses between 2007 and 2013 were analyzed. Clinical and tumor features are shown in **Table 1**. In this study, 62.7% of patients were male, and most were white (83.6%), married (53.5%), and older than 60 years (62.7%). The

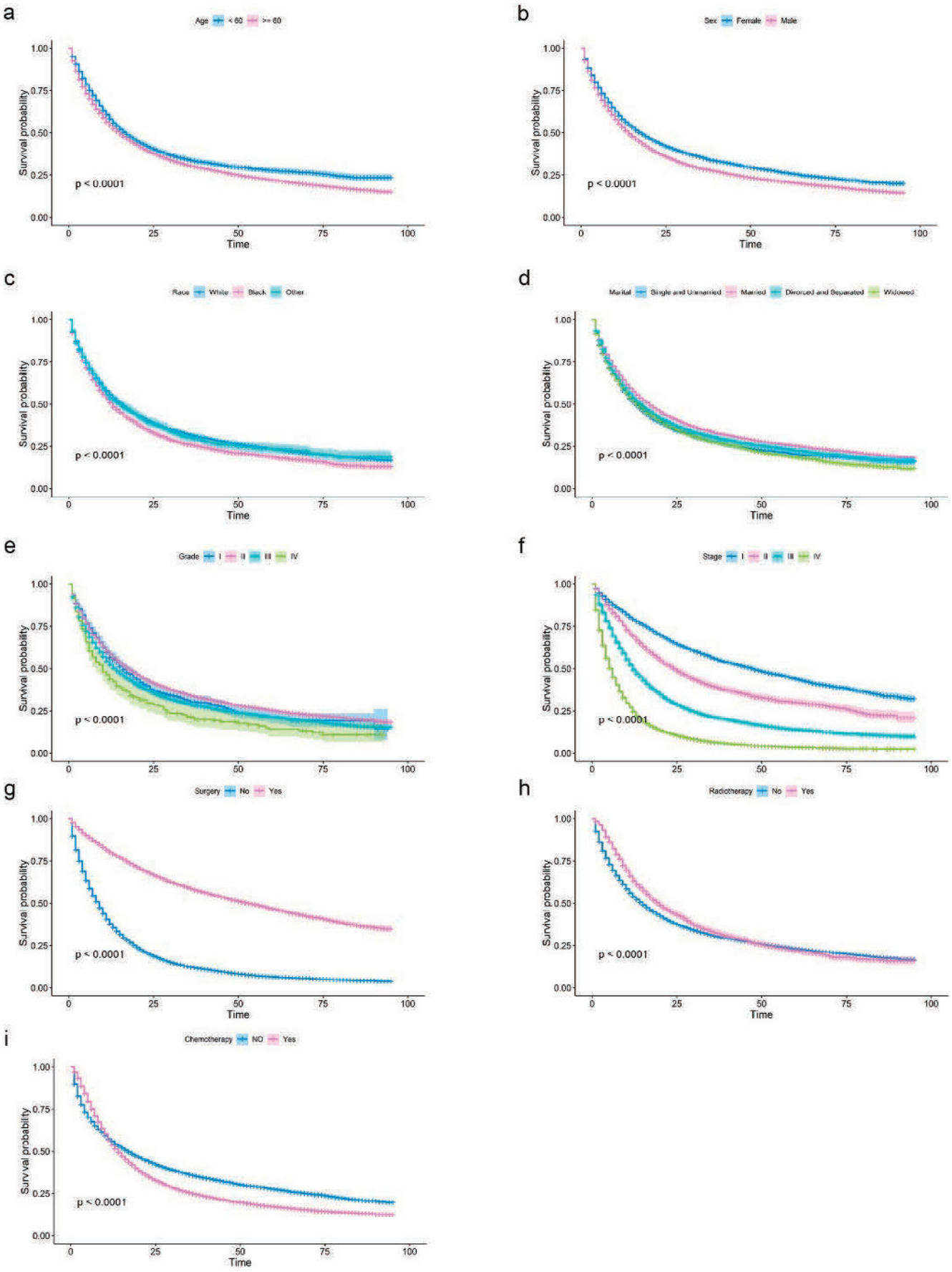

Fig. 1 Overall Kaplan-Meier survival curves for patients in relation to (a) age, (b) sex, (c) race, (d) marital status, (e) tumor differentiation, (f) TNM stage, (g) surgery, (h) radiotherapy, and (i) chemotherapy.

Table 2 Univariate Cox proportional hazards models of overall survival (OS) and cancer-specific survival (CSS)

Variables	OS		CSS	
	HR (95%CI)	*P* value	HR (95%CI)	*p* value
Sex		< 0.001		< 0.001
Female vs. Male	1.18 (1.14-1.21)		1.17 (1.13-1.21)	
Marital status		< 0.001		< 0.001
Single and Unmarried	1.00		1.00	
Married	0.86 (0.82-0.90)		0.83 (0.79-0.87)	
Divorced and Separated	0.93 (0.88-0.99)		0.91 (0.85-0.97)	
Widowed	1.02 (0.97-1.08)		0.95 (0.90-1.01)	
Age, years		< 0.001		0.049
<60 vs. ≥60	1.15 (1.10-1.20)		1.05 (1.00-1.10)	
Race		< 0.001		< 0.001
White	1.00		1.00	
Black	1.16 (1.10-1.21)		1.19 (1.13-1.26)	
Other[a]	1.02 (0.95-1.09)		1.07 (0.99-1.15)	
Grade		< 0.001		< 0.001
I	1.00		1.00	
II	0.93 (0.85-1.02)		0.92 (0.83-1.02)	
III	1.09 (0.99-1.19)		1.12 (1.01-1.24)	
IV	1.34 (1.13-1.58)		1.37 (1.15-1.64)	
TNM stage		< 0.001		< 0.001
I	1.00		1.00	
II	1.51 (1.42-1.60)		1.87 (1.75-2.01)	
III	2.55 (2.45-2.66)		3.38 (3.22-3.55)	
IV	4.94 (4.73-5.15)		6.89 (6.55-7.24)	
Surgery		< 0.001		< 0.001
No vs. Yes	0.26 (0.25-0.27)		0.21 (0.20-0.22)	
Radiotherapy		< 0.001		= 0.001
No vs. Yes	0.87 (0.83-0.92)		0.91 (0.86-0.96)	
Chemotherapy		< 0.001		< 0.001
No vs. Yes	1.18 (1.14-1.21)		1.34 (1.29-1.38)	

[a] Other includes American Indian/Alaska Native, Asian/Pacific Islander, and unknown.

most common tumor grade was poorly differentiated (53.5%), followed by moderately differentiated (42.5%). TNM stage I (33.7%) was the most common stage, followed by stage III (30.8%) and stage IV (25.7%). At the end of follow-up, 16,425 (71.4%) patients had died, including 13,662 (59.4%) who died of LSCC and 2,763 (12.0%) who died of other causes.

Survival Analysis

To identify predictors of survival, data from all 23,004 patients were included in univariate and multivariate Cox regression analyses. As shown in **Figure 1**, survival outcomes differed in relation to age, sex, race, and marital status. In addition, as shown in **Figure 1e, f**, certain clinicopathological factors, such as TNM stage and differentiation, were risk factors affecting survival. As shown in **Figure 1g～i**, surgery, radiotherapy, and chemotherapy were strongly associated with prognosis.

Cox regression analysis was used to further investigate the effects of race, age, sex, marital status, tumor grade, surgery radiotherapy, chemotherapy, and TNM stage (**Table 2, 3**). In univariate analyses, all variables were associated with OS. The independent prognostic variables were race, age, sex, marital status, tumor grade, TNM stage, surgery, radiotherapy, and chemotherapy ($P < 0.05$). After adjusting for other risk factors, the variables identified as independent predictors by a multivariate Cox regression model using stepwise selection were identical to those identified in univariate analyses. In addition, nine variables were independent predictors of CSS ($P < 0.05$).

The developed nomograms, shown in **Figure 2**, are based on important risk factors identified by multivariate analysis of CSS and OS at 3, 5, and 7 years. To calculate 3-, 5-, and 7-year CSS and OS rates, each factor is assigned a point value based on the scales at the top of the nomogram, and the total points are summed. Then, the 3-, 5-, and 7-year CSS and OS rates are obtained by using

Table 3 Multivariate Cox proportional hazards models of overall survival (OS) and cancer-specific survival (CSS)

Variables	OS		CSS	
	HR (95%CI)	*P* value	HR (95%CI)	*p* value
Sex		< 0.001		< 0.001
Female vs. Male	1.21 (1.17-1.25)		1.17 (1.12-1.21)	
Marital status		< 0.001		< 0.001
Single and Unmarried	1.00		1.00	
Married	0.93 (0.88-0.97)		0.92 (0.87-0.97)	
Divorced and Separated	1.03 (0.98-1.10)		1.02 (0.96-1.09)	
Widowed	1.07 (1.01-1.13)		1.02 (0.96-1.09)	
Age, years		< 0.001		< 0.001
<60 vs. ≥60	1.18 (1.13-1.24)		1.12 (1.07-1.17)	
Race		= 0.001		0.005
White	1.00		1.00	
Black	0.92 (0.88-0.97)		0.92 (0.87-0.97)	
Other[a]	0.93 (0.86-1.00)		0.95 (0.88-1.03)	
Grade		< 0.001		< 0.001
I	1.00		1.00	
II	1.01 (0.92-1.10)		0.99 (0.90-1.10)	
III	1.06 (0.96-1.16)		1.07 (0.96-1.18)	
IV	1.32 (1.12-1.56)		1.32 (1.10-1.58)	
TNM stage		< 0.001		< 0.001
I	1.00		1.00	
II	1.77 (1.66-1.88)		2.14 (1.99-2.29)	
III	2.05 (1.95-2.15)		2.53 (2.40-2.68)	
IV	3.43 (3.26-3.61)		4.44 (4.19-4.70)	
Surgery		< 0.001		
No vs. Yes	0.36 (0.34-0.37)		0.32 (0.31-0.34)	< 0.001
Radiotherapy				
No vs. Yes	1.08 (1.03-1.14)	= 0.003	1.11 (1.05-1.18)	< 0.001
Chemotherapy				
No vs. Yes	0.63 (0.61-0.65)	< 0.001	0.66 (0.63-0.68)	< 0.001

[a] Other includes American Indian/Alaska Native, Asian/Pacific Islander, and unknown.

the point scale at the bottom of the nomogram. The models showed good accuracy: the C-index values were 0.732 (95% CI = 0.728-0.736) for the OS model and 0.754 (95% CI = 0.750-0.758) for the CSS model, which indicate relatively good model discrimination in predicting 3-, 5-, and 7-year CSS and OS rates for LSCC patients. The calibration curve based on bootstrap resampling validation is shown in **Figure 3, 4**. The points are close to the 45-degree line; therefore, the nomograms are well standardized.

Discussion

In this study, we analyzed data from 23,004 cases in the SEER database to establish and validate a prognostic nomograms for predicting CSS and OS incidence at 3, 5, and 7 years for LSCC. The results of internal bootstrap resampling to evaluate discriminant performance of the nomograms showed that they fit well with actual observations in predicting 3-, 5-, and 7-year CSS and OS rates, according to the C-index.

Because of the high incidence, poor prognosis, recurrence, and metastasis of LSCC, it remains a substantial public health burden worldwide; however, no unified, efficient model to predict survival has been developed for international use. There is thus an urgent need for better strategies to identify patients at high risk for poor survival. Here, we developed two nomograms that utilize meaningful prognostic factors that are readily available in daily clinical practice.

The nomograms provide a simple graph to represent complex statistical models that quantify personal risk factors, and are potentially useful for broad application in clinical practice and research[20]. Liang et al.[21] developed a nomogram based on the results of an analysis of 6,111 patients with resected non-small-cell lung cancer in China. Variables used in the nomogram included sex,

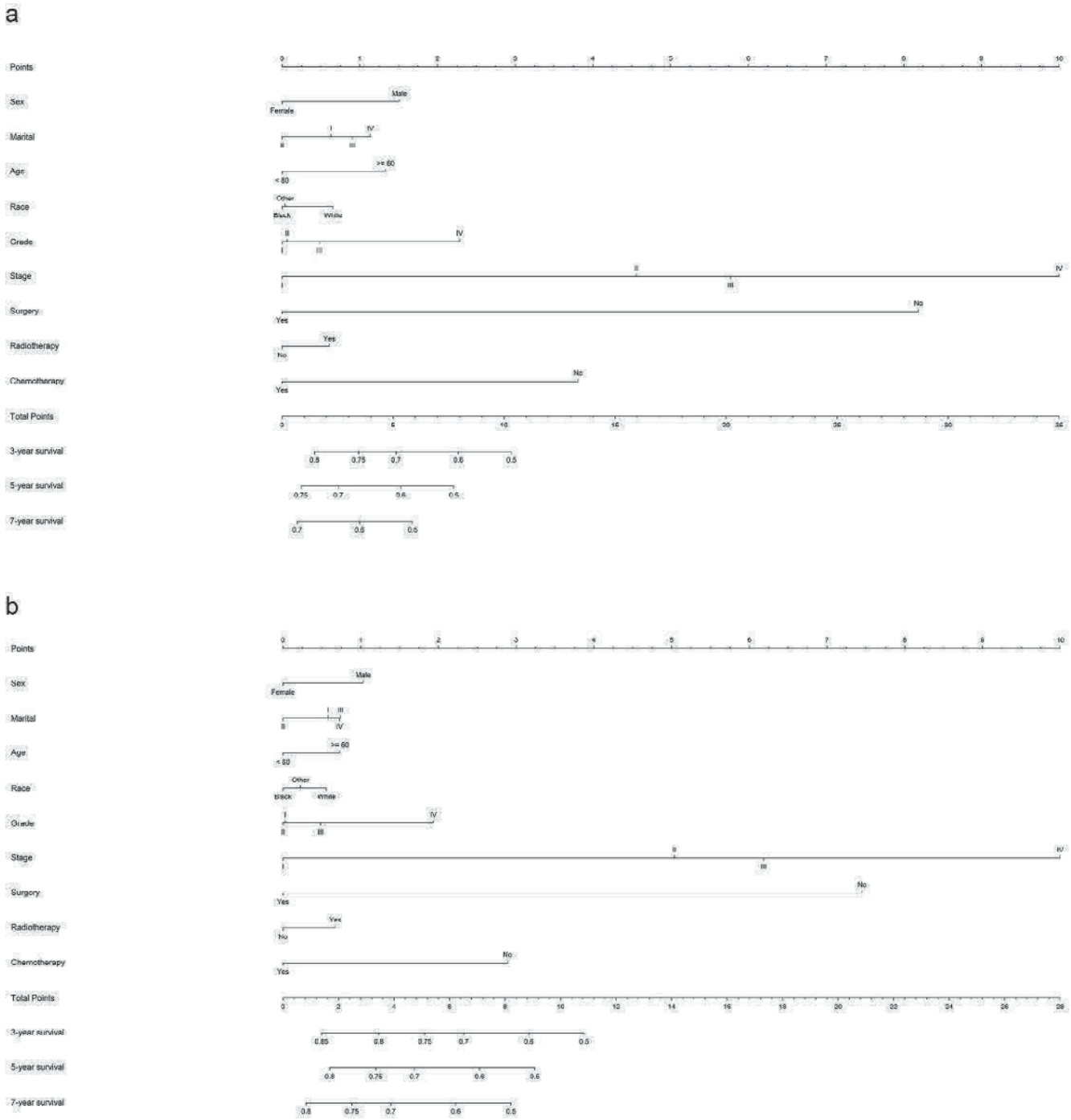

Fig. 2 Nomograms for predicting 3-, 5-, and 7-year (a) overall survival (OS) and (b) cancer-specific survival (CSS) of patients with squamous cell lung cancer.

age, histological status, sampled lymph nodes, T stage, and N stage; the C-index, 0.71, was similar to those for our nomograms. Although several nomograms have been constructed in order to predict the survival rate of patients with lung cancer, prediction accuracy is not entirely satisfactory. Wang et al.[22] established a predictive nomogram for OS in patients with locally advanced lung squamous cell carcinoma, including serum tumor markers such as CEA, CYFRA 21-1, overall stage, and KPS (Karnofsky performance status); however, the accuracy of OS prediction for the nomogram was low, and the C-index was only 0.62. In the present study, nomograms were established for patients with LSCC and showed high compatibility, with C-index values of 0.754 for CSS and 0.732 for OS. In addition, the good prediction quality of the calibration plots ensured the reliability and repeatability of the nomograms. The present nomograms will help in clinical pre-treatment identification of patients

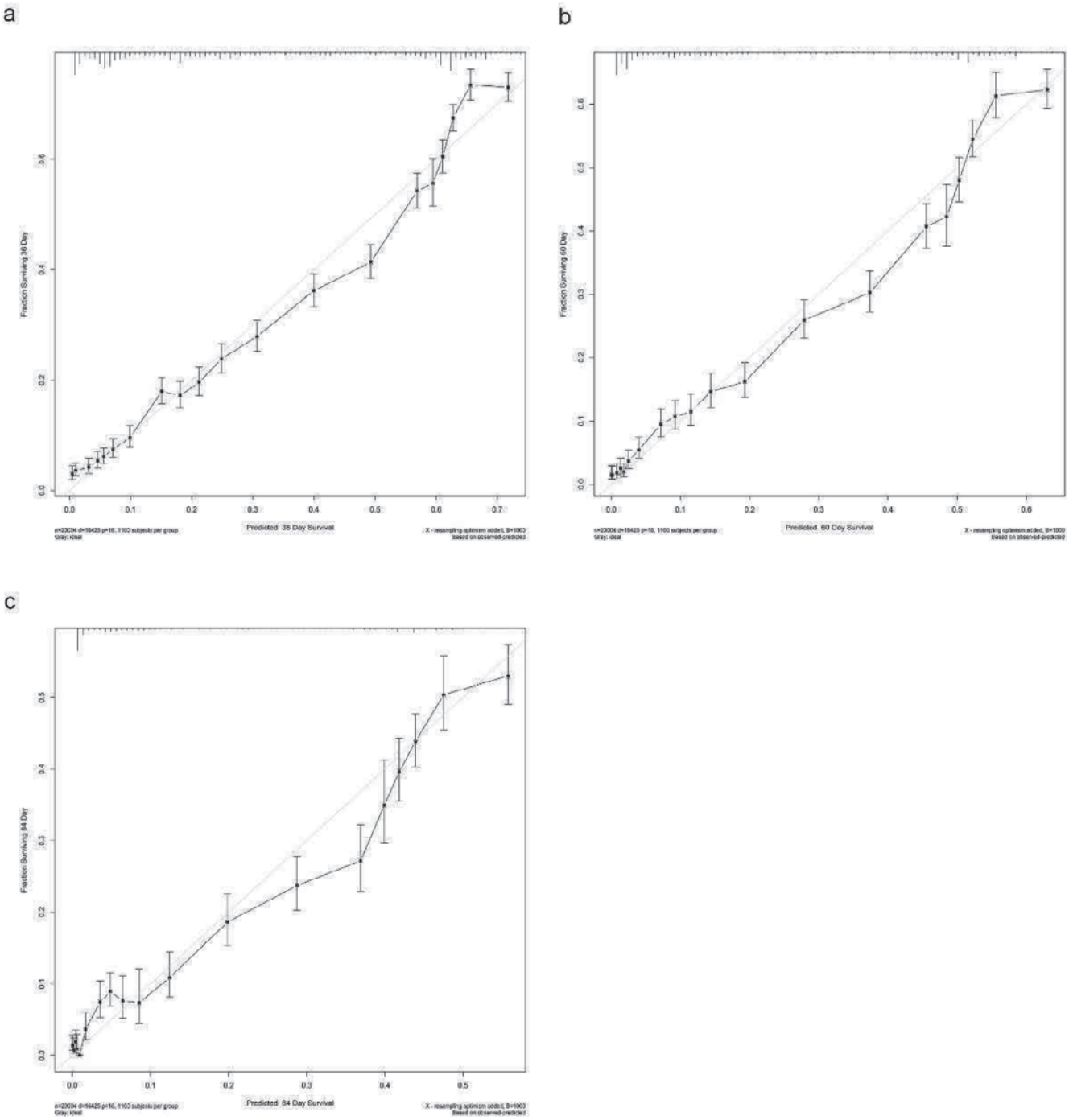

Fig. 3 Calibration plots of nomogram for predicting 3-, 5-, and 7-year overall survival (a, b, c). The X-axis represents the nomogram-predicted probability of survival; the Y-axis represents the actual OS probability.

with a low chance of survival, thus ensuring better clinical decisions and personalized patient follow-up.

The nomogram is a simple graph that represents complex statistical models that quantify personal risk factors, and is potentially broadly applicable in clinical practice and research[23]. Use of nomograms to predict survival is simple. First, a vertical line is drawn from each clinical variable to the "points" line in the nomogram. Then, all the "points" are summed to calculate the "total points", and a vertical line is drawn from "total points" to the "CSS" and "OS" lines to obtain the corresponding survival values. For example, for a 60-year-old (1.3 points) female patient (0 points) with grade IV disease (2.3 points), the nomograms yield 3-, 5-, and 7-year OS values of 74%, 67%, and 61%, respectively. The present nomograms will enable pre-treatment identification of patients with a potentially low survival rate, thereby improving clinical decision-making and follow-up.

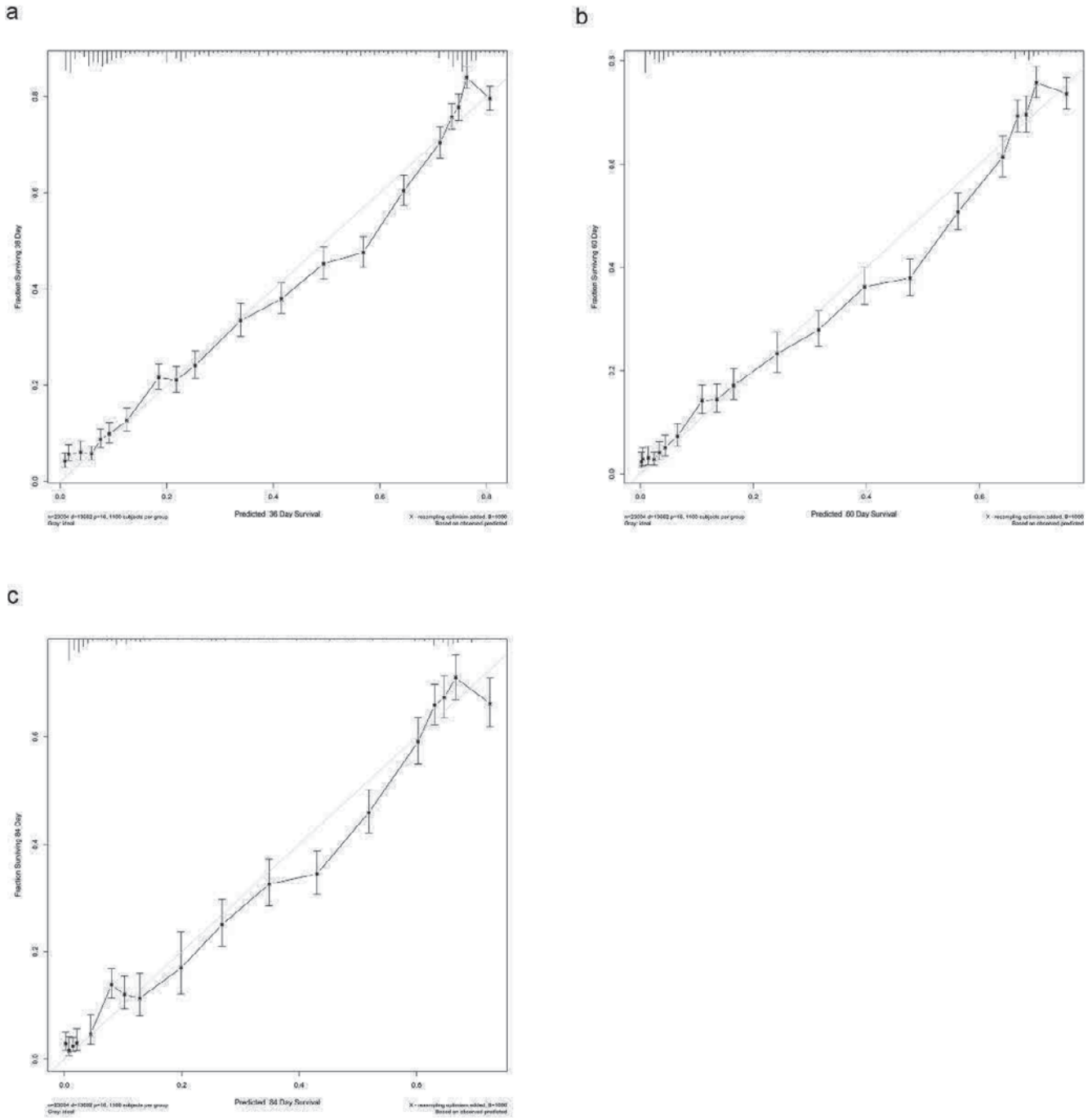

Fig. 4 Calibration plots of nomogram for predicting 3-, 5-, and 7-year cancer-specific survival (CSS) (a, b, c). The X-axis represents the nomogram-predicted probability of survival; the Y-axis represents the actual CSS probability.

The main advantages of this study are its large scale and the simplicity of the models. The SEER database provides rich, detailed samples with which to explore risk factors, ensuring the development of an accurate predictive model. A total of 23,004 eligible patients were included in this study, and statistics for large samples from population-based cancer registries are broader and more reliable than those from single-center studies. In addition, the model variables are easily obtained, and the predictive hazard assessment of patients with LSCC was more complete. Furthermore, our nomograms have good power in predicting CSS and OS, and the presentation of the nomograms was confirmed by calibration. Using the scoring system, we were able to monitor LSCC prognosis more effectively and select treatment strategies to improve survival time.

Limitations

Like all previously published studies that use the various SEER databases, this study shares the known limitations

of using large population-based datasets. First, some misclassification of patient information, such as tumor size and subtype, is inevitable, although regular audits were conducted. Second, because of the retrospective nature of the SEER database, selection bias is possible. Third, the SEER dataset does not provide data for some important clinicopathological variables related to prognosis, such as vascular invasion, smoking, and performance status; future studies should include these factors. Finally, because nomograms do not include all prognostic factors and cannot always provide accurate prognoses in clinical practice, predictive values from nomograms are only for clinician reference.

Conclusions

In summary, using data from an enormous population, we developed and evaluated nomograms for forecasting 3-, 5-, and 7-year CSS and OS that exhibited good prognostic performance in a cohort of patients with LSCC.

Acknowledgements: We are grateful to all the study participants. RJZ and XLG carried out the study design, analysis, and interpretation of data and drafted the manuscript. MLH, MMW, and HQX conceived the study, participated in its design and coordination, and helped draft the manuscript. All authors read and approved the final manuscript.

Conflict of Interest: The authors declare no competing interests.

References

1. Janssen-Heijnen ML, Coebergh JW: The changing epidemiology of lung cancer in Europe. Lung Cancer 2003; 41: 245–258.
2. Locher C, Debieuvre D, Coëtmeur D, Goupil F, Molinier O, Collon T, Dayen C, Le Treut J, Asselain B, Martin F, Blanchon F, Grivaux M: Major changes in lung cancer over the last ten years in France: the KBP-CPHG studies. Lung Cancer 2013; 81: 32–38.
3. B'chir F, Laouani A, Ksibi S, Arnaud MJ, Saguem S: Cigarette filter and the incidence of lung adenocarcinoma among Tunisian population. Lung Cancer 2007; 57: 26–33.
4. Jemal A, Miller KD, Ma J, Siegel RL, Fedewa SA, Islami F, Devesa SS, Thun MJ: Higher Lung Cancer Incidence in Young Women Than Young Men in the United States. N Engl J Med 2018; 378: 1999–2009.
5. Malvezzi M, Carioli G, Bertuccio P, Boffetta P, Levi F, La Vecchia C, Negri E: European cancer mortality predictions for the year 2017, with focus on lung cancer. Ann Oncol 2017; 28: 1117–1123.
6. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A: Global cancer statistics, 2012. CA Cancer J Clin 2015; 65: 87–108.
7. Zhao Y, Qiao W, Wang X, Yin H, Cui J, Cui Y, Chen X, Hu J, Lu H, Meng Q, Wang Y, Cai L: 14-3-3ζ/TGFβR1 promotes tumor metastasis in lung squamous cell carcinoma. Oncotarget 2016; 7: 82972–82984.
8. Dong YH, Lin JW, Wu LC, Chen CY, Chang CH, Chen KY, Lai MS: Examining the association between statins and lung cancer incidence in patients with type 2 diabetes mellitus. J Formos Med Assoc 2014; 113: 940–948.
9. Yan LZ, Dressler EV, Adams VR: Association of hypertension and treatment outcomes in advanced stage non-small cell lung cancer patients treated with bevacizumab or non-bevacizumab containing regimens. J Oncol Pharm Pract 2018; 24: 209–217.
10. Carreras-Torres R, Johansson M, Haycock PC: Obesity, metabolic factors and risk of different histological types of lung cancer: A Mendelian randomization study. PLoS One 2017; 12: e0177875.
11. Costa G, Thuler LC, Ferreira CG: Epidemiological changes in the histological subtypes of 35,018 non-small-cell lung cancer cases in Brazil. Lung Cancer 2016; 97: 66–72.
12. Fang C, Wang W, Feng X, Sun J, Zhang Y, Zeng Y, Wang J, Chen H, Cai M, Lin J, Chen M, Chen Y, Li Y, Li S, Chen J, Zhou Z: Nomogram individually predicts the overall survival of patients with gastroenteropancreatic neuroendocrine neoplasms. Br J Cancer 2017; 117: 1544–1550.
13. Touijer K, Scardino PT: Nomograms for staging, prognosis, and predicting treatment outcomes. Cancer 2009; 115 (13 Suppl): 3107–3111.
14. Surveillance Research Program: National Cancer Institute SEER*Stat software (www.seer.cancer.gov/seerstat) version 8.3.5.
15. Warren JL, Klabunde CN, Schrag D, Bach PB, Riley GF: Overview of the SEER-Medicare data: content, research applications, and generalizability to the United States elderly population. Med Care 2002; 40: IV-3–IV-18.
16. Abdel-Rahman O: Validation of the prognostic value of new sub-stages within the AJCC 8th edition of non-small cell lung cancer. Clin Transl Oncol 2017; 19: 1414–1420.
17. Team RC: R: A language and environment for statistical computing. Vienna, Austria. R Foundation for Statistical Computing; 2014.2014.
18. Wolbers M, Koller MT, Witteman JC, Steyerberg EW: Prognostic models with competing risks: methods and application to coronary risk prediction. Epidemiology 2009; 20: 555–561.
19. Huitzil-Melendez FD, Capanu M, O'Reilly EM, Duffy A, Gansukh B, Saltz LL, Abou-Alfa GK: Advanced hepatocellular carcinoma: which staging systems best predict prognosis? J Clin Oncol 2010; 28: 2889–2895.
20. Bianco FJ Jr: Nomograms and medicine. Eur Urol 2006; 50: 884–886.
21. Liang W, Zhang L, Jiang G, Wang Q, Liu L, Liu D, Wang Z, Zhu Z, Deng Q, Xiong X, Shao W, Shi X, He J: Development and Validation of a Nomogram for Predicting Survival in Patients With Resected Non-Small-Cell Lung Cancer. J Clin Oncol 2015; 33: 861–869.
22. Wang J, Jiang W, Zhang T, Wang Q, Liu L, Liu D, Wang Z, Zhu Z, Deng Q, Xiong X, Shao W, Shi X, He J: Increased CYFRA 21-1, CEA and NSE are Prognostic of Poor Outcome for Locally Advanced Squamous Cell Carcinoma in Lung: A Nomogram and Recursive Partitioning Risk Stratification Analysis. Transl Oncol 2018; 11: 999–1006.
23. Bianco FJ Jr: Nomograms and medicine. Eur Urol 2006; 50: 884–886.

(Received, May 1, 2019)
(Accepted, August 23, 2019)