# Validation of a nomogram to predict survival in patients with ESCC based on examined lymph nodes

Peng Liao ( ✉ liaopengvip@163.com )
  First Affiliated Hospital of Nanchang University
Li Cao
  First Affiliated Hospital of Nanchang University
Hang Chen
  First Affiliated Hospital of Nanchang University
Shui-Zi Pang
  First Affiliated Hospital of Nanchang University

---

Research

---

# Abstract

**Background:** Current the number of examined LNs are controversial in predicting the survival of ESCC. We aimed to develop an alternative LN-classification-based nomogram to individualize ESCC prognosis.

**Methods:** Using the data of patients diagnosed with ESCC from SEER database between 2004 and 2015, we determined the cut-off values for the number of LNs examined via the K-adaptive partitioning (KAPS) algorithm. A nomogram predicting the survival of ESCC was performed, internally and externally validated, and evaluated by calibration plot, C-index, and decision curve analysis, and compared to the 7th TNM stage.

**Results:** Totally, we included 3629 patients with detailed information. The optimal cut-off for examined LN number was 8. The C-index for the nomogram was higher than the 7th TNM staging (internal: 0.708; 95%CI, 0.678-0.753 vs 0.601; 95%CI, 0.573-0.656, P<0.001; external: 0.687; 95%CI, 0.601-0.734 vs 0.605; 95%CI, 0.563-0.659, P<0.001). Additionally, the nomogram showed good agreement between internal and external validation. DCA analysis showed no matter in the internal cohort or external cohort, the nomogram showed a greater benefit across the period of follow-up compared to 7th TNM stage.

**Conclusion:** We found examining LNs that was more than 8 benefited for prognosis of patients. Based on these, a nomogram with greater benefit for predicting survival of EC patients than TNM staging was constructed.

# Introduction

Esophageal cancer (EC) is the fifth leading cause of cancer death for men over the world, and esophageal squamous cell carcinoma (ESCC) accounts for a consitituent ratio of 80%-85% among EC cases in Asian countries[1, 2]. Despite improvements in treatments such as surgical resection and adjuvant chemoradiation, the long-term prognosis of ESCC patients remained less than 30%, especically for 5-year survival[3, 4]. Poor outcomes of patients with ESCC are correlated with diagnosis at advanced stage and high propensity for metastasis caused by insufficiency of lymph node dissection (LND) [5, 6]. Moreover, the number of LND remained to be controversial, as the National Comprehensive Cancer Network (NCCN) guidelines recommended, it is necessary to remove at least 12-15 nodes[7, 8]. As 7th edition of the American Joint Committee on Cancer (AJCC) suggested, dissection does benefit as many regional lymph nodes (LNs) as possible[9]. Additionally, according to numerous retrospective studies, it considered an extensive removal of 6−30 LNs to be favourable for survival improvement[10-13]. Meanwhile, other independent factors such as age, grade, and tumor size could also significantly affect the survival [13]. Based on these, we aimed to determined the optimal number of examined LNs by a large-population study.

In our study, the best cut-off points were determined by The K-adaptive partitioning (KAPS) algorithm which was demonstrated as a useful tool to obtain heterogeneous subgroups by survival[14]. Through

the Surveillance, Epidemiology, and End Results (SEER) database, we developed and validated a nomogram based on multivariate analysis results to predict survival of ESCC patients.

# Methods

## Patients

All patients with EC were retrieved from SEER database by National Cancer Institute's SEER*Stat software (version 8.3.6). All patients did not get informed consent because the SEER database was free for public. According to International Classification of Diseases in Oncology (ICD-O-3), tumors with code 8070, 8071, 8072, 8073, 8074, 8075, 8076, and 8078 were identified as squamous cell carcinoma[15, 16]. In our study, patients with ESCC were included according to the following criteria: (1) patients aged more than 20 who were diagnosed as EC by positive histology from 2004 through 2015; (2) patients with a histopathology of squamous carcinoma; (3) patients with detailed information of survival were included; (4) patients with detailed information including Race, Grade, Regional nodes examined, tumor size, T stage, N stage and M stage.

## Clinicopathological factors

The total of clinicopathological variables extracted from the SEER database in our study include age, race, sex, pathology grade, M stage, tumor size, N stage, regional nodes examined, 7th TNM stage and Marital status. The age of patients was divided into three groups, namely <50 years, 50-70 and >70 years. Race was classified into three types: white, black and other. Sex includes male and female. Pathology grade was categorized as well, moderately differentiated type, poorly differentiated and undifferentiated type. Lymph node metastasis (LNM) was described as N1 (Yes), while N0 was negative. M1 (Yes) indicated positive. As for the tumor size, EC were categorized into 4 groups: ≤2 cm, ≤3 cm, ≤5 cm, and >5 cm. With respect to regional nodes examined, according to the result of K-adaptive partitioning (KAPS) algorithm[17], the cut-off value was 8. Therefore, regional nodes examined was divided into two groups: ≤8 and >8. Marital status was recorded as marrried, unmarried and divorced/widowed. In our study, the main observation indicator was cancer-specific survival (CSS). CSS was defined as death attributable to this cancer.

## Statistical analysis

For the basic statistics, patients were divided into two groups according to diagnosed years, and Pearson's Chi square test was utilized to investigate the association among the categorical variables. With respect to CSS for patients with ESCC, we performed curves of survival by survminer package in R software. Furthermore, to analyze the related risk factors for survival, we performed Lasso regression analysis and multivariate cox regression, then constructed nomogram. The internal cohort included patients diagnosed with EC between 2004 and 2009, while the external validation cohort contained participants diagnosed between 2010 and 2015. The predicting value was assessed by C-index value, time-dependent receiver operating characteristic (tdROC) curves and decision curve analysis (DCA)[14, 18,

19]. The nomogram was first internally validated and then externally validated in the independent cohorts. All statistical analysis was figured out via R software (version 3.6.1, StataCorp LLC, College Station,Tex). The main packages used in our study included ggplot2, survival, rms, kaps and survminer package. Chi square test was carried out by SPSS (version 24.0). It was considered to be statistically significant when value of P was less than 0.05.

# Results

## Patients Characteristics

Totally, we enrolled 3629 patients diagnosed as ESCC from 2004 through 2015 from SEER database. As the flow chart showed (Supplementary Figure 1), we firstly identified the diagnosis of ESCC according to to pathological diagnosis, then patients without T stage information (n=20354), N stage information (n=108), M stage information (n=108), survival information (n=8192) and other information (n=4164) were excluded. Baseline characteristics of patients were presented in the Table 1. 1732 patients were diagnosed from 2004 to 2009 and 1897 patients were found from 2010 to 2015. As for the basic characteristics analysis, we found patients with ESCC were more frequent in the patients aged at >50 years older and male patients. In addition, the lymph node metastasis rate is about 54.42%, while patients with distant metastasis accounted for 24.28%. The median survival was 9 months which ranged from 4 to 23 months.

## Grouping of lymph nodes in EC patients

Using KAPS algorithm, we found the optimal cut-off dividing the number of examined lymph nodes (LNs) into two groups was 8, and then performed the Kaplan-Meier survival analysis. As showed in Figure 1A, it could be considered as significantly different between the two groups (P<0.001) for over survival (OS) rate. Also, for analysis of CSS, patients with less than 8 examined LNs had worse prognosis that those who with more than 8 examined LNs (Figure 1B). Additionally, we performed Kaplan-Meier survival analysis in the different stage and found patients with >8 examined  LNs remained to have a better survival (Figure 2). The difference of survival among the two goups for patients in the different TNM stage was statistically significant (P<0.0001).

## EC survival prediction model

Since the variables we extracted were sufficient, to sellect the most suitable characters to predict prognosis, we performed Lasso regression analysis and found age, tumor size, T stage, N stage, M stage and examined LNs were highly associated with survival (Figure 3). Moreover, using multivariate cox analysis model, we identified age, tumor size, T stage, N stage, M stage and examined LNs were independent prognostic factors (Figure 4). Patients with age>=70, tumor size>5cm and advanced stage had poorer prognosis, while patients with examined LNs>8 were associated with better prognosis. Then model of nomogram that predicts survival was performed based on multivariate cox analysis (Figure 5). The nomogram showed T stage contributed the most to prognosis, followed by examined LNs, M stage,

tumor size and age, whereas lymph node metastasis had the least effect for prognosis. As for explannation of nomogram, a straight line can be drawn down at each time point to determine the estimated probability of survival. With respect to each predictor, we could read the points assigned on the 0–10 scale at the top and then add these points. Lastly, find the number on the "Total Points" scale and read the corresponding predictions of 1-, 3-, and 5-year risk.

## Validation and Calibration of Nomogram

In our internal validation, the C-index for the established nomogram to predict CSS (0.708, 95%CI, 0.678-0.753) was significantly higher than that for 7th TNM stage (0.601, 95%, 0.573-0.656, P<0.001) (Table 2). In our external validation, C-index of the new model (0.687, 95%CI, 0.601-0.734) was higher than that of 7th TNM stage (0.605, 95%CI, 0.563-0.659, P<0.001) (Table 2). As indicated by the calibration plots, good agreement was observed between the actual observation and prediction by nomogram for 1-year, 3-year and 5-year CSS (Figure 6A-6C). As for analysis of specificity, the nomogram performed better than TNM staging in both internal cohort (1-year AUC: 0.753 vs 0.653, 3-year AUC: 0.761 vs 0.701, 5-year AUC: 0.783 vs 0.733, P<0.001, Figure 6D-6F) and external cohort (1-year AUC: 0.761 vs 0.641, 3-year AUC: 0.753 vs 0.687, 5-year AUC: 0.75 vs 0.685, P<0.001, Figure 6G-I). Finally to compare the clinical usability between nomogram and TNM stage, we performed DCA and showed results in the Figure 7. For predicting 1-year, 3-year and 5-year CSS, no matter in the internal cohort or external cohort, the nomogram showed a greater benefit across the period of follow-up compared to TNM stage.

# Discussion

Our present investigation shows that the number of examined was demonstrated to be inversely correlated with the survival of ESCC patients, so we figured out the cut-off value of LNs examined. The cut-off value of examined LNs was 8 and we divided patients into two groups by the number of examined LNs: <=8 examined LNs and >8 examined LNs. As we all know, Lasso regression analysis could well avoid confounding factors and made our predicting model more acurate[20]. According to the results (Figure 3), six characters including tumor size, T stage, age, examined LNs, N stage and M stage were identified as the most suitable variables for constructing model. In addition, we performed multivariate regression analysis and the nomogram, finding that nomogram was more accurate in internal and external cohorts than the traditional 7th TNM staging, and showed better clinical usefulness for predicting survival assessed by DCA.

As we all know, examining LNs is an important factor for prognosis of patients with gastrointestinal cancer, and also the association between survival and the number of examined LNs was also demonstrated by previous studies[21, 22]. A large population-based study using the National Cancer Database has showed patients who had more than 15 lymph nodes resected had better survival[23], as well as other studies[24]. In our study, the best cut-off values for examined LNs was 8, which does make a difference for survival evaluation. Moreover, our results were in line with other studies[11, 25, 26]. Published studies showed the area under curves (AUC) of ROC for predicting sensitivity and specifity was

about 0.61, although it was considered as powerful prognostic factor to evaluate survival[27, 28]. Based on the these notion, the nomogram was performed well with an AUC of more than 0.7 compared to 7th TNM staging. For the traditional TNM staging, some studies also thought it was controversial and proposed other different subdivision[11, 29, 30]. With repect to the clinical utility, we performed C-index, tdROC and Calibration plot and demonstrated it was effective predicting model in the external and internal validation. No matter external or internal validation, the value of C-index was about 0.7, suggesting the nomogram model was an effective model[31]. The result of tdROC showed nomogram has a great superiority than 7th TNM staging via comparing the AUC between them[14]. Also, as the Calibration plot suggested, there is a good agreement between the predicting value and actual observations[14]. As for the number of examined LNs in different TNM stage, we found the cut-off values for two groups made a great difference for prognosis. Previous studies reported more than 8 LNs would be beneficial for the improvement of survival[32]. For the Stage IV of EC, Greenstein et al. considered 18 LNs removed were required for accurate staging and improvement of survival[33], while Bollschweiler et al. indicated that 15 examined LNs at least were better for prognosis[34], which greatly depended on the operation and pathological diagnosis[10]. Therefore, this difference of conclusion may be due to the heterogeneity of clinical data.

Our study has some limitations that should be discussed. Firstly, TNM staging was 7th edition rather than 8th edition, which may influence the result of comparion between nomogram and TNM staging system. Hence, it is necessary to perform study to compare the model constructed by our study with 8th TNM staging. Then we excluded many patients who had missing data associated with our collected variables, increasing the selection bias. Next, the variables including examined LNs were sellected depending on each doctor in different clinical centers. Therefore, although this nomogram performed well in the two cohort, it should be applied with great caution when assessing risk of 1-, 3- and 5- year survival. In the future, we will collect our relevant data to incorporate factors above into further research.

# Conclusions

In conclusion, the present study was constructed to investigate the optimal examined LNs. We found examined LNs did benefit for prognosis of patients, which was more favorable as many as 8 examined LNs. Based on the Lasso and multivariate analysis, we developed and validated a nomogram with greater benefit for predicting survival of ESCC patients than TNM staging, which was demonstrated by td-ROC and DCA.

# Abbreviations

EC: Esophageal cancer

ESCC: Esophageal squamous cell carcinoma

LN: lymph node

KAPS：K-adaptive partitioning

SEER：Surveillance, Epidemiology, and End Results

tdROC: time-dependent receiver operating characteristic

DCA: decision curve analysis

CSS: cancer-specific survival

AJCC: American Joint Committee on Cancer

MLR: metastatic lymph node ratio

NCCN: National Comprehensive Cancer Network

# Declarations

## Author contributions

PL : data collection, data analysis, and manuscript writing. LC and HC: data analysis. SZP: project development.

## Conflicts of interest

The authors disclose no conflicts.

## Fundings

Not appliable

## Acknowledgements

Not appliable

## Consent for publication

Not appliable

## Ethics statement

Ethics approval and consent was obtained from SEER database.

## Availability of Data and Materials

Our data was extracted from SEER database.

# References

1. Lin, Y., et al., *Epidemiology of esophageal cancer in Japan and China.* J Epidemiol, 2013. **23**(4): p. 233-42.

2. Bray, F., et al., *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.

3. Napier, K.J., M. Scheerer, and S. Misra, *Esophageal cancer: A Review of epidemiology, pathogenesis, staging workup and treatment modalities.* World J Gastrointest Oncol, 2014. **6**(5): p. 112-20.

4. Pennathur, A., et al., *Oesophageal carcinoma.* Lancet, 2013. **381**(9864): p. 400-12.

5. A, P., et al., *Esophagectomy for T1 esophageal cancer: outcomes in 100 patients and implications for endoscopic therapy.* The Annals of thoracic surgery, 2009. **87**(4): p. 1048-54; discussion 1054-5.

6. A, P., et al., *The "best operation" for esophageal cancer?* The Annals of thoracic surgery, 2010. **89**(6): p. S2163-7.

7. Vashist, Y.K., et al., *Glasgow Prognostic Score is a predictor of perioperative and long-term outcome in patients with only surgically treated esophageal cancer.* Ann Surg Oncol, 2011. **18**(4): p. 1130-8.

8. O'Sullivan B, B.J., John Wiley and Sons Ltd.; Chichester, West Sussex, UK; Hoboken, NJ: , *Oesophagus*, in *UICC Manual of Clinical Oncology*. 2015. p. 280-296.

9. Lin, Z., et al., *Achieving adequate lymph node dissection in treating esophageal squamous cell carcinomas by radical lymphadenectomy: Beyond the scope of numbers of harvested lymph nodes.* Oncol Lett, 2019. **18**(2): p. 1617-1630.

10. Hu, Y., et al., *How does the number of resected lymph nodes influence TNM staging and prognosis for esophageal carcinoma?* Ann Surg Oncol, 2010. **17**(3): p. 784-90.

11. Peyre, C.G., et al., *The number of lymph nodes removed predicts survival in esophageal cancer: an international study on the impact of extent of surgical resection.* Ann Surg, 2008. **248**(4): p. 549-56.

12. Groth, S.S., et al., *Determination of the minimum number of lymph nodes to examine to maximize survival in patients with esophageal carcinoma: data from the Surveillance Epidemiology and End Results database.* J Thorac Cardiovasc Surg, 2010. **139**(3): p. 612-20.

13. Schwarz, R.E. and D.D. Smith, *Clinical impact of lymphadenectomy extent in resectable esophageal cancer.* J Gastrointest Surg, 2007. **11**(11): p. 1384-93; discussion 1393-4.

14. S, W., et al., *A New Metastatic Lymph Node Classification-based Survival Predicting Model in Patients With Small Bowel Adenocarcinoma: A Derivation and Validation Study.* EBioMedicine, 2018. **32**: p. 134-141.

15. Zheng, X.Q., et al., *Incidence, prognostic factors, and a nomogram of lung cancer with bone metastasis at initial diagnosis: a population-based study.* Transl Lung Cancer Res, 2019. **8**(4): p. 367-379.

16. Jayarajah, U., et al., *The incidence and histological patterns of oesophageal cancer in Sri Lanka from 2001 to 2010: Analysis of national cancer registry data.* Eur J Cancer Care (Engl), 2019: p. e13182.

17. S, W., et al., *A New Metastatic Lymph Node Classification-based Survival Predicting Model in Patients With Small Bowel Adenocarcinoma: A Derivation and Validation Study.* EBioMedicine, 2018. **32**(undefined): p. 134-141.

18. JA, H. and M. BJ, *A method of comparing the areas under receiver operating characteristic curves derived from the same cases.* Radiology, 1983. **148**(3): p. 839-43.

19. P, B., D. JF, and J.-G. H, *Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks.* Statistics in medicine, 2013. **32**(30): p. 5381-97.

20. Lu, S., et al., *Development and validation of a nomogram for preoperative prediction of cervical lymph node involvement in thyroid microcarcinoma.* Aging (Albany NY), 2020. **12**.

21. SS, G., et al., *Determination of the minimum number of lymph nodes to examine to maximize survival in patients with esophageal carcinoma: data from the Surveillance Epidemiology and End Results database.* The Journal of thoracic and cardiovascular surgery, 2010. **139**(3): p. 612-20.

22. R, V., et al., *Lymph node evaluation and long-term survival in Stage II and Stage III colon cancer: a national study.* Annals of surgical oncology, 2009. **16**(3): p. 585-93.

23. DN, G., et al., *Comparative Quantitative Lymph Node Assessment in Localized Esophageal Cancer Patients After R0 Resection With and Without Neoadjuvant Chemoradiation Therapy.* Journal of gastrointestinal surgery : official journal of the Society for Surgery of the Alimentary Tract, 2017. **21**(9): p. 1377-1384.

24. Samson, P., et al., *Extent of Lymphadenectomy Is Associated With Improved Overall Survival After Esophagectomy With or Without Induction Therapy.* Ann Thorac Surg, 2017. **103**(2): p. 406-415.

25. XB, D.J., *Clinical implication of the innovations of the 8 edition of the TNM classification for esophageal and esophago-gastric cancer.* Journal of thoracic disease, 2018. **10**: p. S2671-S2681.

26. Bogoevski, D., et al., *Is it time for a new TNM classification in esophageal carcinoma?* Ann Surg, 2008. **247**(4): p. 633-41.

27. Strong, V.E., et al., *Impact of the 7th Edition AJCC staging classification on the NCCN clinical practice guidelines in oncology for gastric and esophageal cancers.* J Natl Compr Canc Netw, 2013. **11**(1): p. 60-6.

28. Gulben, K., et al., *Prognostic significance of number of lymph node metastasis on survival in patients with pathological T3 esophageal carcinoma.* Neoplasma, 2017. **64**(1): p. 131-135.

29. Akutsu, Y. and H. Matsubara, *The significance of lymph node status as a prognostic factor for esophageal cancer.* Surg Today, 2011. **41**(9): p. 1190-5.

30. Wang, N., et al., *Prognostic significance of lymph node ratio in esophageal cancer.* Tumour Biol, 2015. **36**(4): p. 2335-41.

31. JM, A., et al., *Nomogram to predict the benefit of radiation for older patients with breast cancer treated with conservative surgery.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2012. **30**(23): p. 2837-43.

32. Z, L., et al., *Achieving adequate lymph node dissection in treating esophageal squamous cell carcinomas by radical lymphadenectomy: Beyond the scope of numbers of harvested lymph nodes.* Oncology letters, 2019. **18**(2): p. 1617-1630.

33. AJ, G., et al., *Effect of the number of lymph nodes sampled on postoperative survival of lymph node-negative esophageal cancer.* Cancer, 2008. **112**(6): p. 1239-46.

34. E, B., et al., *Staging of esophageal carcinoma: length of tumor and number of involved regional lymph nodes. Are these independent prognostic factors?* Journal of surgical oncology, 2006. **94**(5): p. 355-63.

# Tables

**Table 1: Patients' demographics, clinical characteristics at diagnosis**

| Variables | Total (%) | 2004-2009 | 2010-2015 | P Value |
|---|---|---|---|---|
| n | 3629 | 1732 | 1897 | |
| Age | | | | 0.415 |
| <50 | 226 (6.23%) | 115 (6.64%) | 111 (5.85%) | |
| 50-70 | 2066 (56.93%) | 969 (55.95%) | 1097 (57.82%) | |
| >70 | 1337 (36.84%) | 648 (37.41%) | 689 (36.32%) | |
| Race | | | | 0.993 |
| White | 731 (67.13%) | 1085 (62.64%) | 1192 (62.83%) | |
| Black | 220 (20.2%) | 441 (25.46%) | 481 (25.36%) | |
| Other | 138 (12.67%) | 206 (11.9%) | 224 (11.8%) | |
| Sex | | | | 0.536 |
| Male | 2349 (64.73%) | 1130 (64.57%) | 1219 (64.26%) | |
| Female | 1280 (35.27%) | 602 (35.43%) | 678 (35.74%) | |
| Pathology Grade | | | | 0.052 |
| Well differentiated | 207 (5.7%) | 83 (4.79%) | 124 (6.54%) | |
| Moderately differentiated | 1875 (51.67%) | 880 (50.81%) | 995 (52.45%) | |
| Poorly differentiated | 1517 (41.8%) | 754 (43.53%) | 763 (40.22%) | |
| Undifferentiated | 30 (0.83%) | 15 (0.87%) | 15 (0.79%) | |
| Lymph node Metastasis | | | | 0.000 |
| NO | 1654 (45.58%) | 848 (48.96%) | 806 (42.49%) | |
| Yes | 1975 (54.42%) | 884 (51.04%) | 1091 (57.51%) | |
| Metastasis | | | | 0.845 |
| No | 2748 (75.72%) | 1309 (75.58%) | 1439 (75.86%) | |
| Yes | 881 (24.28%) | 423 (24.42%) | 458 (24.14%) | |
| Tumor size | | | | 0.172 |
| ≤2cm | 403 (11.1%) | 190 (10.97%) | 213 (11.22%) | |
| ≤3cm | 486 (13.4%) | 234 (13.51%) | 252 (13.28%) | |
| ≤5cm | 1205 (33.2%) | 604 (34.87%) | 601 (31.68%) | |
| >5cm | 1535 (42.3%) | 704 (40.65%) | 831 (43.81%) | |
| Regional_nodes_examined | | | | 0.156 |
| <=8 | 3052 (84.1%) | 1441 (83.2%) | 1611 (84.92%) | |
| >8 | 577 (15.9%) | 291 (16.8%) | 286 (15.08%) | |
| Marital_status | | | | 0.078 |
| Married | 1756 (48.39%) | 845 (48.79%) | 911 (48.02%) | |
| Unmarried | 1065 (29.35%) | 528 (30.48%) | 537 (28.31%) | |
| Divorced/Widowed | 808 (22.26%) | 359 (20.73%) | 449 (23.67%) | |
| T stage | | | | 0.00 |
| T1 | 422 (11.63%) | 187 (5.15%) | 235 (12.39%) | |
| T2 | 1428 (39.35%) | 623 (35.97%) | 805 (43.49%) | |
| T3 | 1024 (28.21%) | 526 (30.37%) | 498 (26.25%) | |
| T4 | 755 (20.8%) | 396 (22.86%) | 359 (18.92%) | |
| 7th TNM stage | | | | 0.051 |
| I | 520 (14.33%) | 276 (13.91%) | 244 (12.86%) | |
| II | 1033 (28.47%) | 486 (41.89%) | 547 (28.83%) | |
| III | 1200 (33.07%) | 552 (35.43%) | 648 (34.16%) | |
| IV | 881 (24.28%) | 423 (8.77%) | 458 (24.14%) | |
| Median survival (M) | 9 (4-23) | 9 (4-23) | 9 (4-22) | |

Table 2: Accuracy of the prediction score of the nomogram and TNM stage for estimating prognosis of patients with ESCC

| Variable | Value (95%CI) | |
| --- | --- | --- |
| | Internal validation | External validation |
| C index for nomogram | 0.708 (0.678-0.753) | 0.687 (0.601-0.734) |
| C index for TNM stage | 0.601(0.573-0.656) | 0.605(0.563-0.659) |
| 1 year AUC for nomogram | 0.753(0.711-0.821) | 0.761(0.715-0.831) |
| 3 year AUC for nomogram | 0.761(0.712-0.813) | 0.753 (0.659-0.818) |
| 5 year AUC for nomogram | 0.783(0.753-0.848) | 0.75 (0.753-0.847) |
| 1 year AUC for TNM stage | 0.653(0.611-0.701) | 0.641(0.605-0.715) |
| 3 year AUC for TNM stage | 0.701(0.675-0.784) | 0.687 (0.655-0.738) |
| 5 year AUC for TNM stage | 0.733(0.613-0.781) | 0.685 (0.643-0.727) |

# Figures
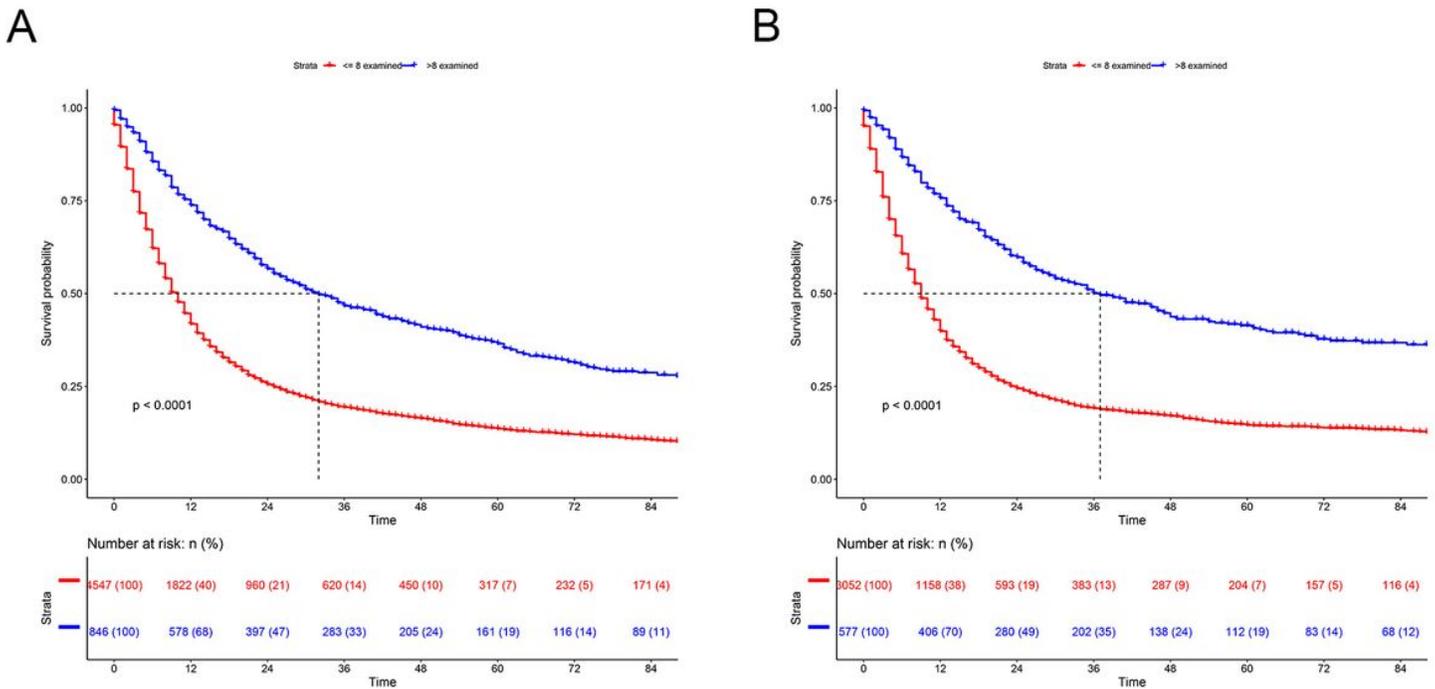


## Figure 1

Kaplan-Meier survival analysis according to the number of examined lymph nodes. (A) OS , (B) CSS
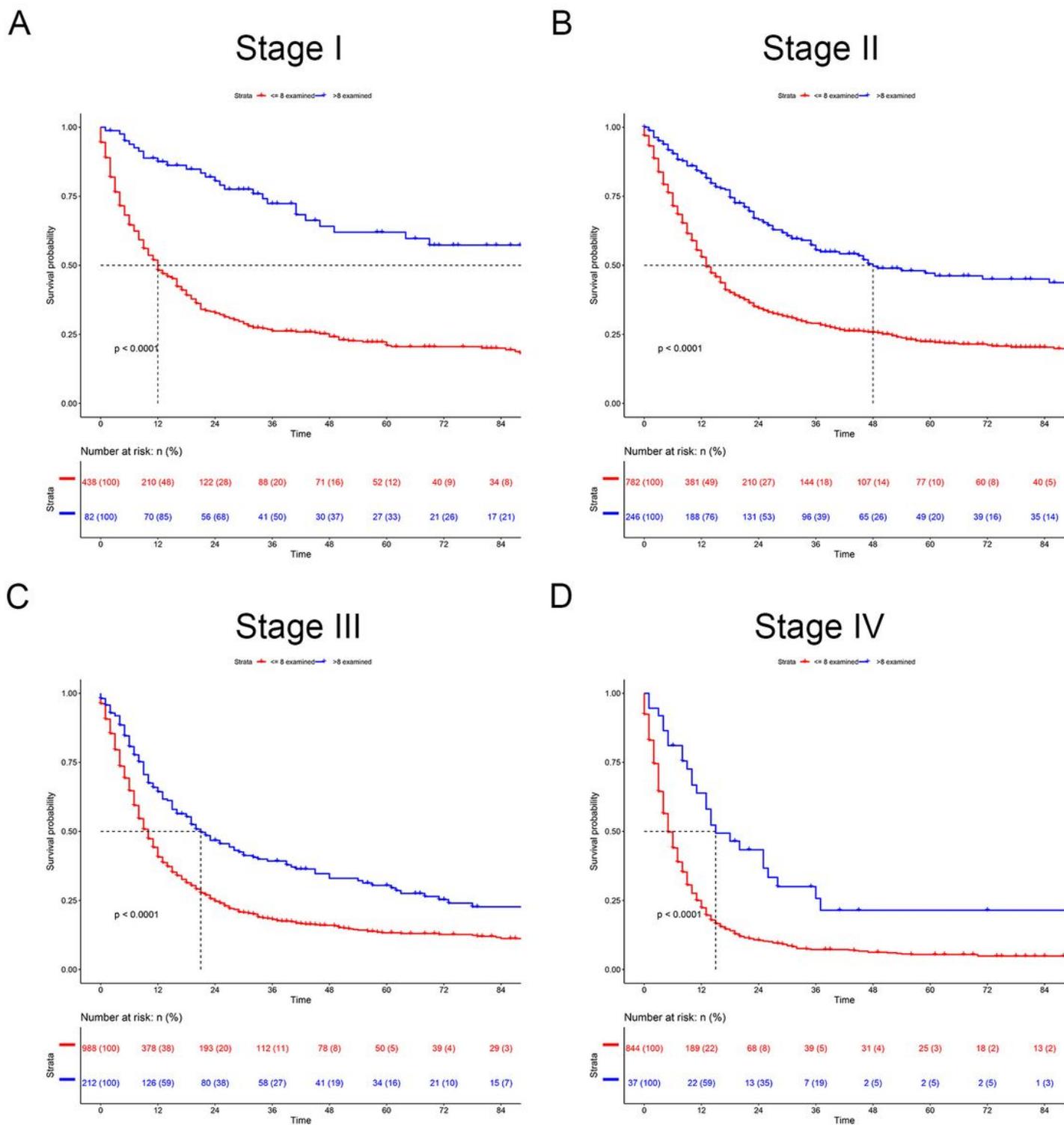
**Figure 2**

Kaplan-Meier survival analysis for CSS according to the number of examined lymph nodes in different TNM stage. (A) Stage I, (B) Stage II, (C) Stage III, (D) Stage IV
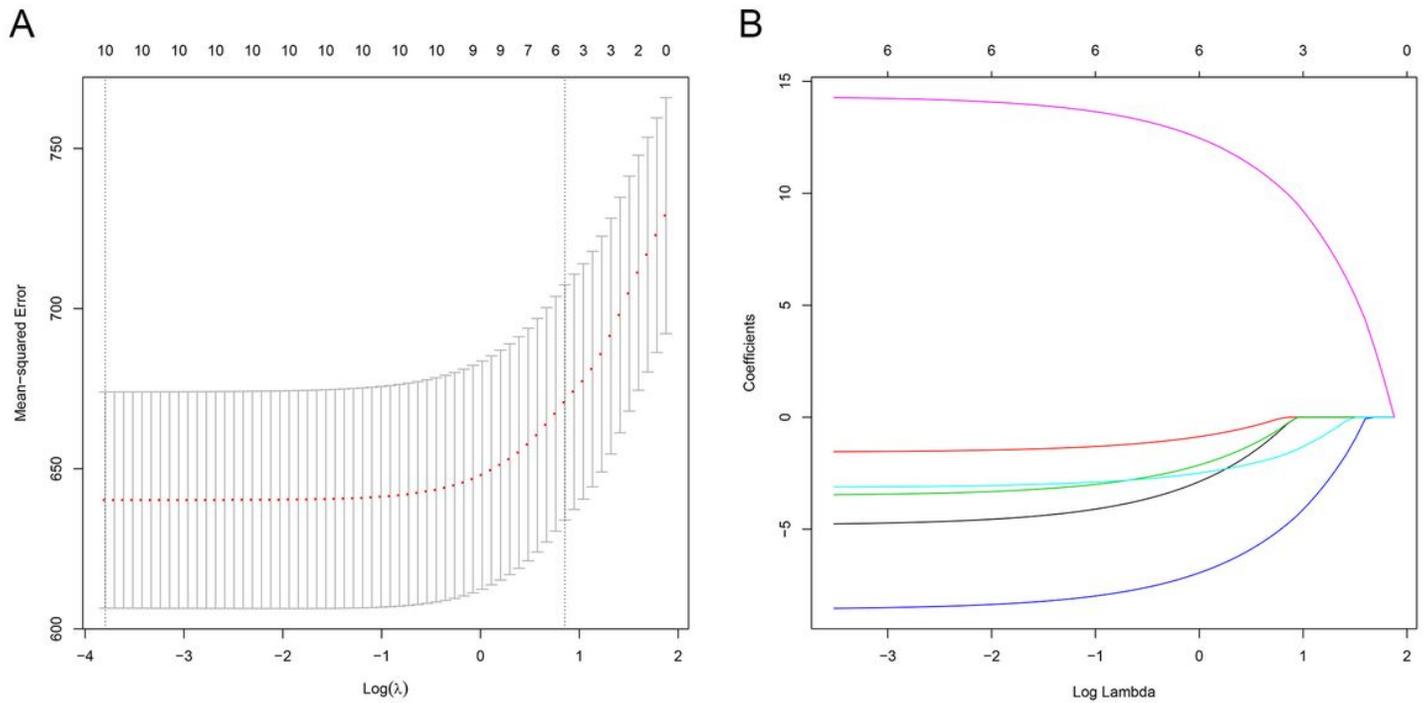
**Figure 3**

The Association between our extrected clinical characteristics of patients and survival was analysed by Lasso regression analysis. Final predictors include age, T stage, N stage, M stage, tumor size and examined LNs. (A) Suitable parameter ($\lambda$) selection in the LASSO model used 5-fold cross-validation via minimum criteria. We plotted the partial likelihood deviance (binomial deviance) curve versus log ($\lambda$). 2 dotted vertical lines were drawn at the optimal values applying the minimum criteria and the 1 standard error of the minimum criteria (the 1-SE criteria). (B) LASSO coefficient profiles of the 6 variables. We produced a coefficient profile plot against the log ($\lambda$) sequence. A suitable $\lambda$ was chosen when log ($\lambda$)= 0.82 and resulted in 6 variables. LASSO=least absolute shrinkage and selection operator, SE=standard error.

## Hazard ratio of ESCC

| | | | | |
|---|---|---|---|---|
| **Age** | <50 (N=226) | reference | | |
| | 50-69 (N=2066) | 1.059 (0.896 – 1.251) | | 0.5032 |
| | >=70 (N=1337) | 1.444 (1.217 – 1.713) | | <0.001 *** |
| **T_stage** | T1 (N=422) | reference | | |
| | T2 (N=1428) | 1.189 (1.032 – 1.370) | | 0.0169 * |
| | T3 (N=1024) | 1.440 (1.245 – 1.665) | | <0.001 *** |
| | T4 (N=755) | 2.094 (1.799 – 2.436) | | <0.001 *** |
| **N_stage** | N0 (N=1654) | reference | | |
| | N1 (N=1975) | 1.103 (1.017 – 1.197) | | 0.0179 * |
| **M_stage** | M0 (N=2748) | reference | | |
| | M1 (N=881) | 1.701 (1.557 – 1.858) | | <0.001 *** |
| **tumor_size** | <=2cm (N=403) | reference | | |
| | <=3cm (N=486) | 1.216 (1.029 – 1.437) | | 0.0215 * |
| | <=5cm (N=1205) | 1.292 (1.117 – 1.495) | | <0.001 *** |
| | >5 (N=1535) | 1.700 (1.473 – 1.963) | | <0.001 *** |
| **Regional_nodes_examined** | <=8 (N=3052) | reference | | |
| | >8 (N=577) | 0.513 (0.454 – 0.581) | | <0.001 *** |

# Events: 2654; Global p-value (Log-Rank): 1.5489e-157
AIC: 39043.52; Concordance Index: 0.67
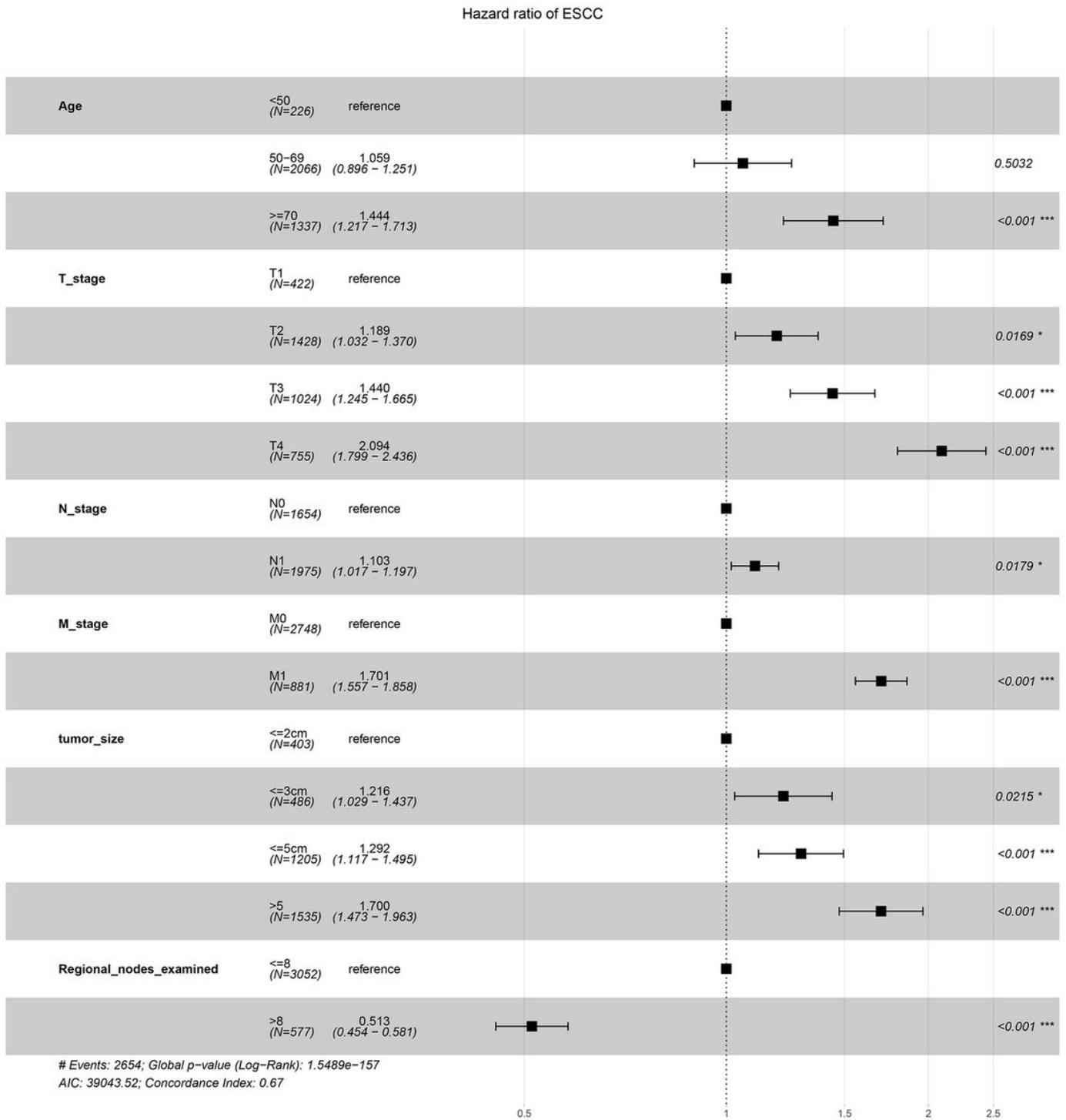
0.5     1     1.5     2     2.5

## Figure 4

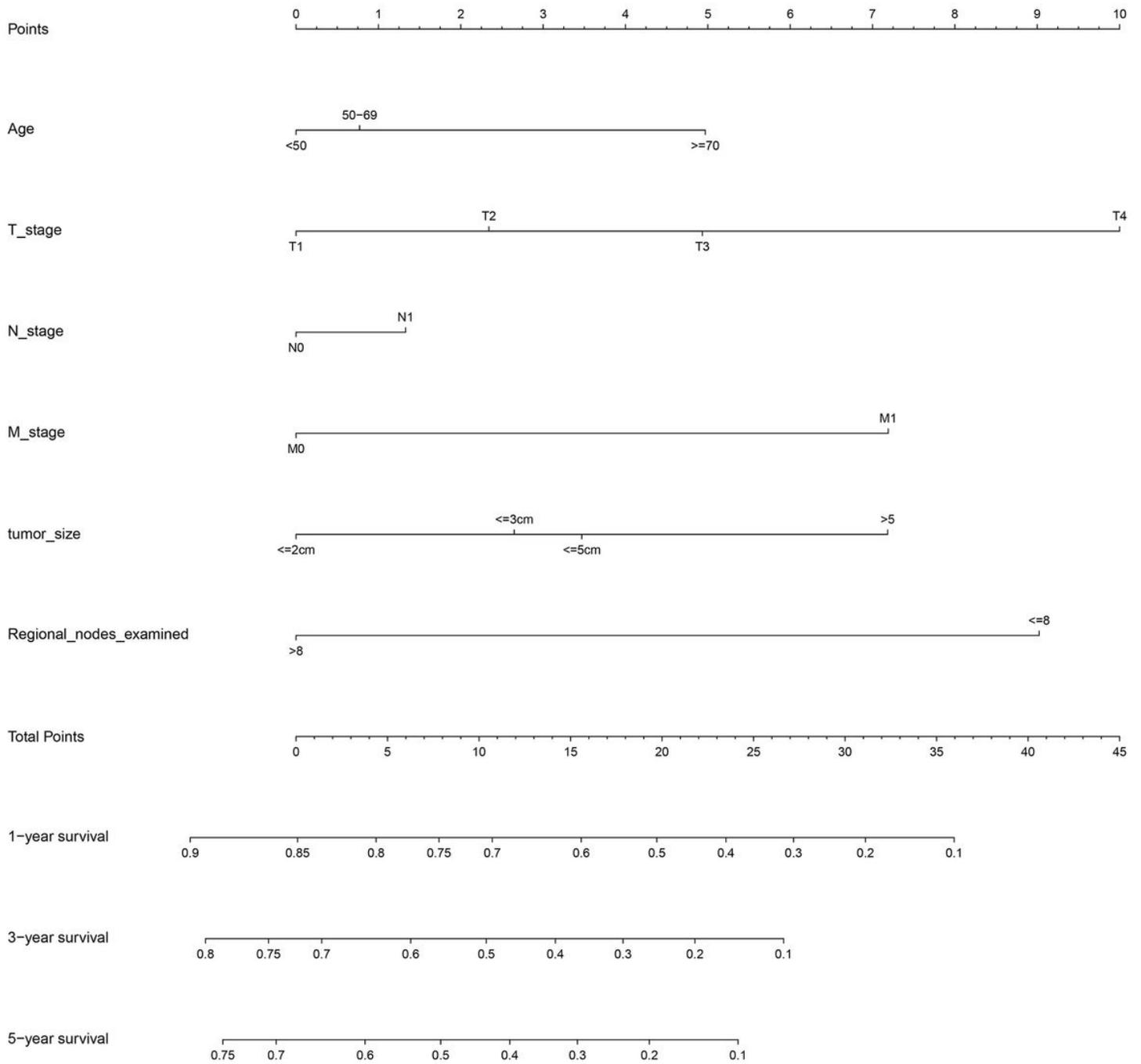Multivariate cox regression analysis was performed to identify the risk factors for survival.

## Figure 5

Nomogram predicted 1-, 3- and 5-year cancer specific survival for patients with resected ESCC.
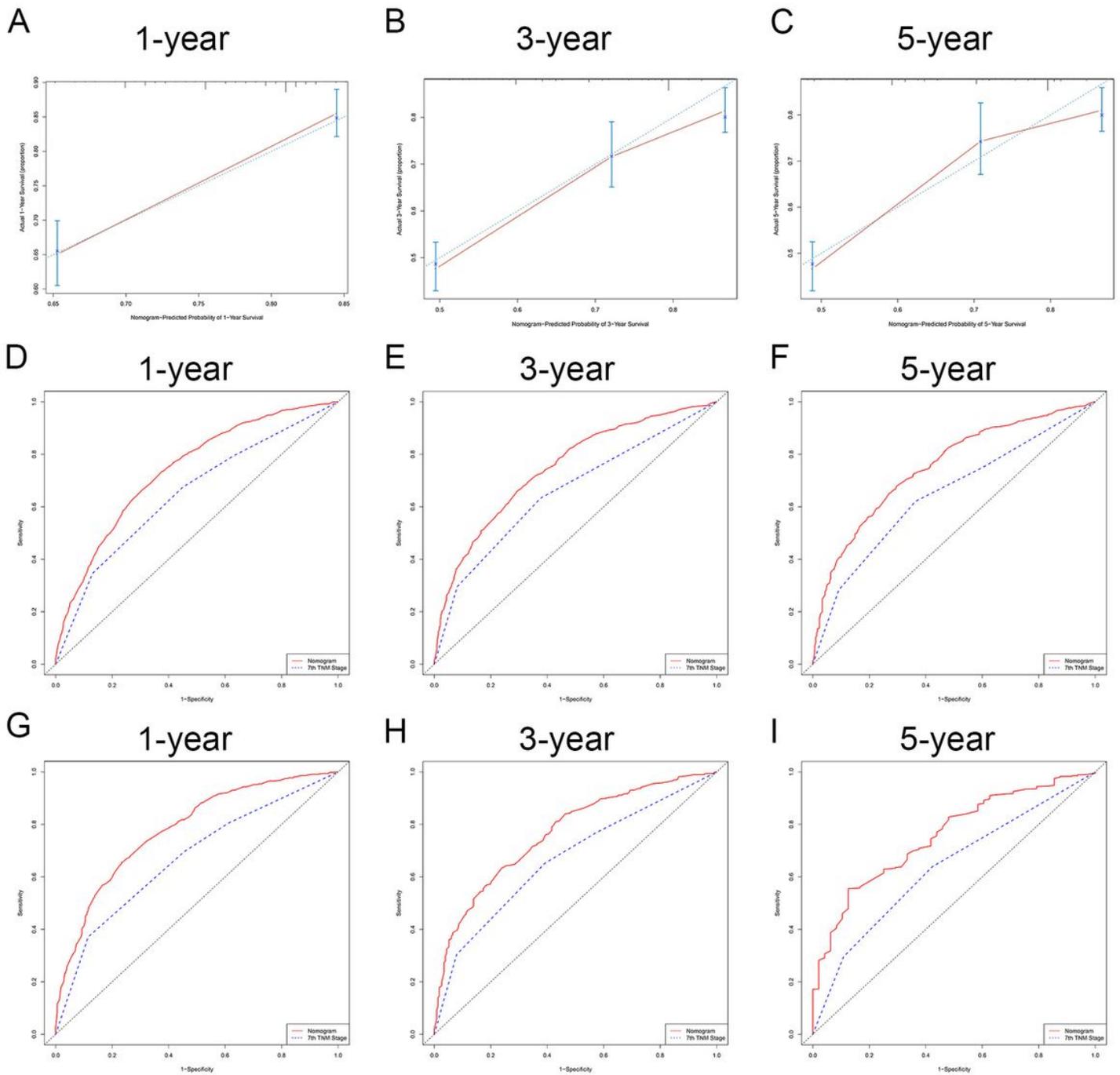
**Figure 6**

The calibration curves at 1, 3 and 5 year point in the 2004–2015 cohort and ROC curve of the Nomogram and 7th TNM Stage in prediction of prognosis of patients from 2004 to 2015. (A) 1-year survival. (B) 3-year survival. (C) 5-year survival. Nomogram-predicted CSS is plotted on the x-axis; cancer specific survival is plotted on the y-axis. (D-F) ROC curve for 1-, 3- and 5- year point in 2004-2009 cohort. (G-I) ROC curve for 1-, 3- and 5- year point in 2010-2015 cohort.
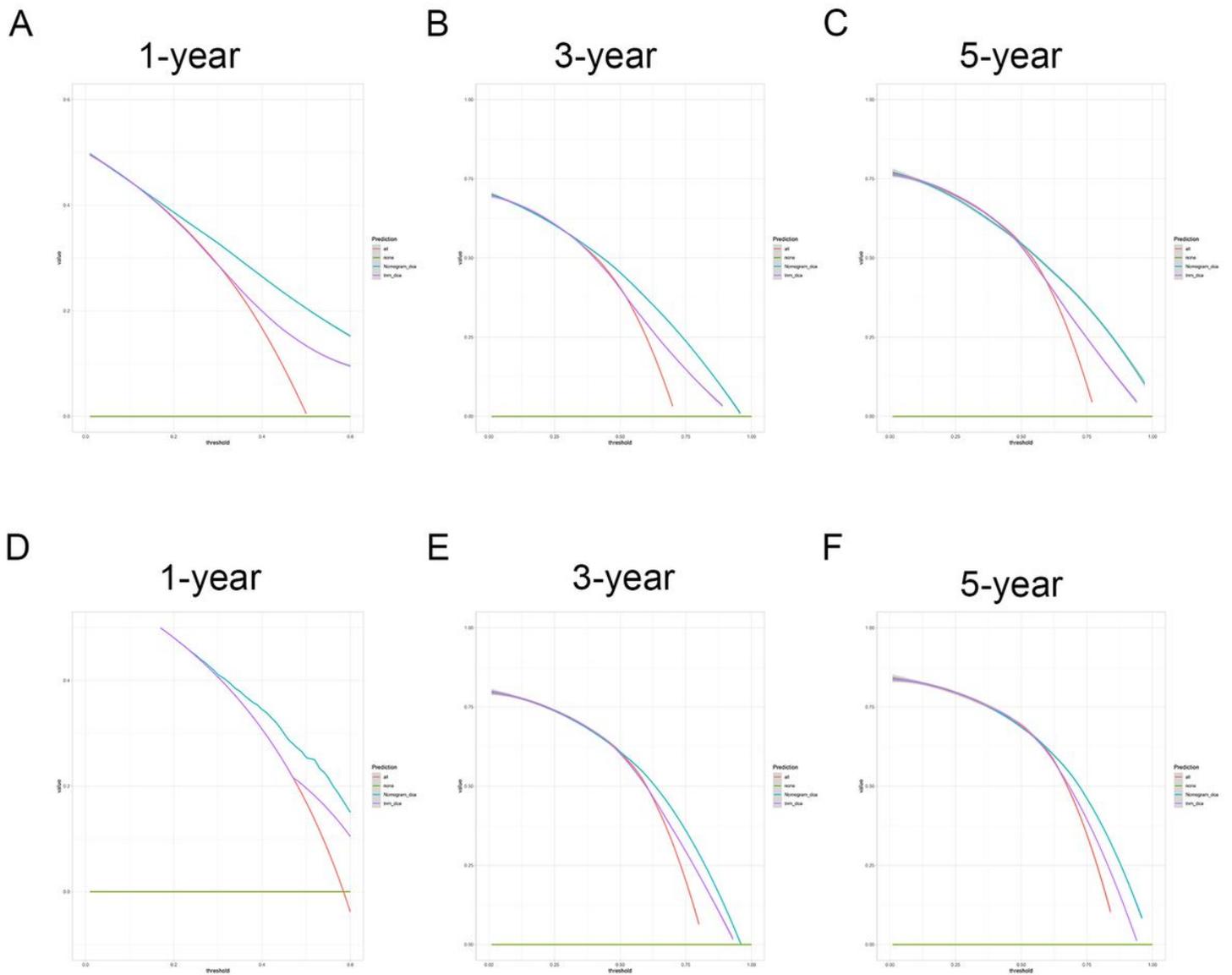
**Figure 7**

Decision curve analysis for the Nomogram and the model 7th TNM Stage in prediction of prognosis of patients. (A-C) 1-, 3- and 5- year point in 2004-2009 cohort. (D-F) 1-, 3- and 5- year point in 2010-2015 cohort.

# Supplementary Files

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

- SupplementaryFigure1.jpg