

# Development and Analysis of a Predictive Nomogram Assessing Cancer Risk in a Chinese Cohort of Patients Presenting with Pulmonary Nodules

**Zufang Liao**

Li Huili Hospital Affiliated to Ningbo University

**Rongjiong Zheng**

Ningbo Yinzhou NO.2 Hospital

**Ni Li**

Li Huili Hospital, Ningbo University

**Guofeng Shao** (✉ [sgf1958@sina.com](mailto:sgf1958@sina.com))

Li Huili Hospital, Ningbo University

---

## Research Article

**Keywords:** Lung cancer, pulmonary nodules, predictors, nomogram

**Posted Date:** December 1st, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-1119406/v1>

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

[Read Full License](#)

---

# Abstract

**Background:** Lung cancer is a major global threat to public health for which a novel prognostic nomogram is urgently needed.

**Patients and methods:** Here, we designed a novel prognostic nomogram using a training dataset consisting of 178 pulmonary nodules for design and 124 nodules for external validation. The R 'caret' package was used to separate patients for design into two groups, including a training cohort (n=126) for model construction and an internal validation cohort (n=52). Optimal feature selection for this model was achieved using the least absolute shrinkage and selection operator regression (LASSO) model. C-index values, calibration plots, and decision curve analyses were used to gauge the discrimination, calibration, and clinical utility, respectively, of this predictive model. Validation was then performed with the validation cohort.

**Results:** A predictive nomogram was successfully constructed incorporating hypertension status, plasma fibrinogen levels, serum uric acid (SUA) levels, triglyceride (TG) and high-density lipoprotein (HDL) levels, density, spicule sign, ground-glass opacity (GGO), and pulmonary nodule size. This model exhibited good discriminative ability, with a C-index value of 0.795 (95% CI: 0.720–0.870), and was well-calibrated. When we used the validation cohort to evaluate the model, the C-indexes were 0.886 (95% CI: 0.800–0.972) and 0.817 (95% CI: 0.747–0.897) for internal validation and external validation, respectively. Decision curve analyses indicated the clinical value of this predictive nomogram when used at a lung cancer possibility threshold of 9%.

**Conclusion:** The nomogram constructed in this study, which incorporates hypertension status, plasma fibrinogen levels, SUA, TG, HDL, density, spicule sign, GGO status, and pulmonary nodule size was able to reliably predict lung cancer risk in this Chinese cohort of patients presenting with pulmonary nodules.

## Introduction

Lung cancer is a form of malignancy arising from the unrestrained growth of bronchial and lung cells<sup>1-2</sup>, and it is one of the leading causes of mortality in the world<sup>3</sup>. Rates of lung cancer have been rising rapidly in recent years, particularly in more heavily industrialized nations<sup>4</sup>. Currently, lung cancer patients exhibit 5-year survival rates of approximately 16.6%<sup>5</sup>, and approximately 1 million individuals in China are forecast to suffer from lung cancer by the year 2025, with this nation exhibiting the highest global incidence of lung cancer.

Key risk factors associated with lung cancer incidence include specific genetic mutations, smoking, and environmental exposures such as air pollution. There is also some evidence suggesting that factors such as a poor diet, alcohol intake, estrogen levels, the smoking of marijuana, and infection with human papillomavirus (HPV), human immunodeficiency virus (HIV), and Epstein-Barr virus may increase lung cancer risk, although such evidence remains somewhat inconclusive<sup>6</sup>. Analyses of patient computed

tomography (CT) scans often reveal pulmonary nodules in patients, and many models have been developed to gauge the link between such nodules and lung cancer risk, including the Brock model<sup>7</sup>. These models, however, often do not take epidemiological variables, clinical findings, and CT scan results into consideration at the same time, limiting their value as predictors of the relative risk of a given pulmonary nodule being malignant. The development of more reliable and accurate predictive tools has the potential to enable early intervention and treatment for lung cancer patients, maximizing their odds of positive outcomes. Herein, we analyzed 26 variables with potential relevance to the diagnosis of a given pulmonary nodule as being benign or malignant based on previous studies<sup>7,8,9,10</sup>.

By analyzing epidemiological, clinical, and CT-related factors for patients with pulmonary nodules that had undergone surgical treatment, we sought to develop a simple but robust predictive model that would enable the relative assessment of lung cancer risk based only on characteristics that can be readily assessed prior to surgery or other therapeutic interventions.

## **Materials And Methods**

### **Patients**

The Ethics Committee of the affiliated Lihuili Hospital of Ningbo University, Lihuili hospital approved this study (approval no KY2020PJ141). Enrolled patients were individuals from China recruited at the Lihuili Hospital of Ningbo University of Medicine between October 2020 to April 2021 and the later external validation cohort from June 2021 to October 2021. Eligible patients were individuals that had undergone resective surgeries following pulmonary nodule identification. Patients provided written informed consent to participate in this study. Any patients diagnosed with serious cognitive or physical impairments, or other serious diseases were excluded from the study cohort. Data including patient clinical, demographic, and disease-related characteristics were retrieved from patient medical records. The R 'caret' package was used to separate patients for design into two groups at a 7:3 ratio at random, including a training cohort for nomogram construction and an internal validation cohort.

### **Statistical analysis**

Data are given as numbers (percentages), and were analyzed using R (v 4.1.2, <https://www.R-project.org>). The LASSO method, which enables the reduction of high-dimensional datasets, was utilized as a means of selecting the optimal predictors of lung cancer risk among included pulmonary nodule patients. Those features that yielded non-zero coefficient values in this LASSO regression analysis were retained for nomogram incorporation. The final predictive model was constructed via a multivariate logistic regression approach, with all significance levels being two-sided. A training cohort was used to develop the predictive model, with calibration curves being used to assess nomogram calibration. Significant calibration curve results were indicative of a model that was not perfectly calibrated. Model discrimination performance was assessed based on the value of Harrell's C-index. Validation of this nomogram was additionally performed to calculate an accurate C-index value. Decision curve analyses

were used to assess the clinical utility of this lung cancer risk nomogram by quantifying the net benefit at different probability thresholds in the lung cancer cohort, with the net benefit being calculated by subtracting the proportion of patients with false-positive results from the proportion of patients with true-positive results and by assessing the relative harm of failing to intervene as compared to the potential negative outcomes associated with an unnecessary intervention. Receiver operating characteristic (ROC) curves were also used to assess the precision of this predictive risk model.

## Results

### Patient characteristics

In total, data from 178 patients for design with pulmonary nodules that visited our clinic between October 2020 and April 2021 were included in this study, as well as the 124 patients from June 2021 to October 2021 for external validation. Patients for design were randomly assigned to training and validation cohorts at a 7:3 ratio. Patients in the training cohort (74 males, 52 females, mean age: 60.64±10.50 years [range: 31–86 years]) were separated into groups with benign nodules and malignant lesions. For details regarding the demographic and clinical characteristics of patients in these groups, see Table 1.

### Feature selection and predictive model development

In total, 26 potentially relevant features were evaluated for inclusion in a predictive model. Of these features, 9 were ultimately selected through a LASSO regression analysis of the 126 patients in the training cohort (Figure 1A and B). These features included hypertension status, serum uric acid (SUA), triglyceride (TG) and high-density lipoprotein (HDL) levels, plasma fibrinogen levels, density, ground-glass opacity (GGO) status, spicule sign, and pulmonary nodule size. A predictive model incorporating these seven variables was developed using the training cohort (Figure 2). Nodule density was defined as being “low” when it exhibited a CT value that was higher than that of pulmonary tissue but lower than that of blood, “intermediate” for nodules with solid and GGO components, and “high” when CT values were greater for the nodule than for blood.

### Assessment of predictive risk model performance

Calibration curves for this predictive nomogram when used to analyze the training cohort revealed it to be well-calibrated, with a C-index value of 0.795 (95% CI: 0.720–0.870) (Figure 3A). Similarly, the C-indexes for internal validation and external validation were 0.886 (95% CI: 0.800–0.972) (Figure 3B) and 0.817 (95% CI: 0.747–0.897) (Figure 3C) respectively, consistent with the discriminative value of this model, suggesting that it exhibits good predictive capabilities.

### Analysis of model clinical utility

Decision curve analyses for this predictive nomogram were next performed (Figure 4). These analyses revealed that at a threshold probability of a patient and a doctor is >9 and <98%, respectively, then this

nomogram exhibits value as a means of predicting lung cancer risk. Net benefit was comparable with some overlaps within this range when assessing lung cancer risk.

## ROC curve analysis

ROC curve analyses of the three cohorts confirmed the predictive value of this risk model, with an area under the curve value of 0.795 for training cohort (Figure 5 A), 0.886 for internal validation cohort (Figure 5 A) and 0.817 for external validation cohort (Figure 5 C).

## Discussion

Nomograms are valuable predictive tools that have been widely utilized in oncology and other clinical and research fields, offering a user-friendly approach to intuitively assessing the odds of a given prognosis or outcome based on a set of specific variables, thereby aiding in clinical decision-making<sup>11</sup>. Many models for the treatment of pulmonary nodules were established on the basis of certain epidemiological variables and CT scan results. However, clinical findings such as blood biomarkers are also very important for the diagnosis of lung cancer<sup>12</sup>. Moreover, for some of these variables, such as GGO, the surgical criteria are not well defined such that treatments are often conducted according to surgeons' own experience<sup>13,14</sup>. As such, we herein sought to develop a new nomogram capable of predicting the relative risk of malignancy when evaluating patients with pulmonary nodules.

Here, we designed and validated a novel predictive model capable of assessing the risk of a given lung nodule being benign or malignant based on analysis of data from patients that had undergone pulmonary nodule resection. The resultant model incorporated demographic, disease-, and treatment-related features to easily predict the odds of a given pulmonary nodule corresponding to a lung cancer diagnosis. The model developed herein was accurate, and exhibited good calibration and discrimination in our validation cohort. The C-index value in this validation cohort was also high, indicating that the nomogram can be accurately used to gauge patient risk of pulmonary nodule malignancy<sup>11</sup>.

Prior studies have confirmed that hypertension is a common comorbidity in cancer patients<sup>15</sup>. Several mechanisms may explain this observation, including the fact that hypertension can increase VEGF levels in the plasma<sup>16</sup>. We identified hypertension as a risk factor for lung nodule malignancy. Fibrinogen has also been significantly linked to the risk of lung cancer in the past<sup>17</sup>, with Kuang et al. having demonstrated that a combination of the beta and gamma chains of fibrinogen may offer value as a sensitive biomarker for differentiating between lung nodules that are benign and malignant<sup>18</sup>, potentially explaining the significance of plasma fibrinogen levels in our model. Dovell et al. demonstrated that higher SUA levels were associated with a somewhat higher risk of overall cancer incidence, including lung cancer<sup>19</sup>. Xie et al. also found hyperuricemia to be linked to increased rates of cancer and associated mortality<sup>20</sup>, potentially explaining the inclusion of SUA in our predictive risk model. HDL levels have been

shown to be negatively correlated with the risk of lung cancer in one cohort study<sup>21</sup>, with other studies having similarly supported the existence of lower HDL levels in lung cancer patients relative to healthy individuals<sup>22</sup>. HDL levels are also readily measured in a clinical context. TG levels have also been reported to be positively correlated with lung cancer incidence in an analysis conducted by Lin et al. of 4673 lung cancer patients in a cohort of 685,852 individuals<sup>18</sup>. Low and high TG levels have been linked to higher rates of lung cancer in a prospective cohort study<sup>23</sup>. With respect to spicule sign, Fang et al. previously conducted a case-control study demonstrating that stage I lung adenocarcinoma patients were significantly more likely to exhibit this finding<sup>24</sup>. GGO findings have been reported to be associated with cancer rates as high as 63%, with many surgeons believing that GGO nodules should be resected, particularly if they grow in size. Persistent GGO nodules may be indicative of a greater risk of malignancy when solid components are evident<sup>25</sup>. Tu et al. found CT density to be a valuable feature when differentiating between nodules that were malignant and benign<sup>26</sup>. Qiu et al. further determined that solitary ground-glass opacity nodule size and density upon high-resolution T evaluation were associated with invasive adenocarcinoma risk<sup>27</sup>. Nodule size may be the most important variable included in our predictive module, given that nodule diameter is a key determinant of treatment under the British Thoracic Society guidelines<sup>28</sup> and Fleischner Society Guidelines<sup>29</sup>. For nodules  $\geq 10$  mm in diameter, the odds of malignancy in the NELSON screening study were 15.2%<sup>30</sup>. As such, we included nodule diameter as the size variable in the present study.

Herein, we found that pulmonary nodules  $> 8$  mm in size were more likely to be malignant than smaller nodules (57.53% vs. 40.63%), suggesting that a predictive model including this parameter, after being appropriately calibrated, may aid in improving lung cancer patient outcomes by providing individualized predictions of risk. Herein, we thus developed a risk nomogram that may aid clinicians in differentiating between patients with benign or malignant lung nodules. It may also aid in the optimal selection of pulmonary nodules in the context of clinical research. For example, this model might be used to aid investigators in selecting patients with larger nodules and other risk-related findings when selecting subjects for surgical procedures or other interventions. Early interventions including CT scans, biochemical analyses of blood samples, and family support can better benefit low-risk patients, while regular clinical examination can ensure the appropriate monitoring of lung nodules to better guide the appropriate assessment of patient prognosis.

Accurate prognostic evaluations can aid surgeons in predicting lung cancer risk in individual patients, ensuring timely intervention for high-risk patients while reducing the need for interventional treatment in low-risk patients. Accurately predicting the risk of lung cancer in a given patient is very challenging, and appropriate measurements together with multifaceted interventional approaches are thus the most reliable approach to detecting and evaluating patients with pulmonary nodules. Further research on this topic is warranted as the accurate detection of pulmonary nodules alone is necessary but insufficient for treating affected patients, underscoring directions for future study.

## Limitations

There are multiple limitations to this study. For one, the sample size of this study was limited, and all patients were enrolled from a single center over a relatively limited study period. However, nomograms established by Chen et al.<sup>31</sup> and Luo et al.<sup>32</sup>, with the training and validation cohorts (61/101 and 32/43 patients, respectively), exhibited good accuracy. Additionally, risk factor analyses did not incorporate all possible risk factors that may be relevant to the differentiation between benign and malignant nodules. Other relevant factors not included in this analysis included the number of nodules and specific comorbidity incidence rates. Lastly, while a bootstrap testing approach was used to validate our nomogram, the patients used for this validation approach may not be sufficient to ensure the generalizability of these data to patients from other countries or regions. As such, further external validation in a wider pulmonary nodule patient population will be essential in the future.

## Conclusions

In summary, we herein designed a novel nomogram with good accuracy that offers value as a means of differentiating between benign and malignant pulmonary nodules, enabling clinicians to better plan patient treatment. Such individualized risk analyses offer clinicians an opportunity to appropriately monitor and treat patients. However, further work will be needed to validate this nomogram in larger patient populations and to establish whether the treatment decisions made based on this nomogram will reduce rates of incorrect diagnosis and treatment planning for patients with pulmonary nodules.

## Abbreviations

LASSO: least absolute shrinkage and selection operator, SUA serum uric acid TG triglyceride HDL high-density lipoprotein GGO ground-glass opacity HPV human papillomavirus, HIV: human immunodeficiency virus, CT: computed tomography, ROC: Receiver operating characteristic, VEGF: vascular endothelial growth factor, SE: standard error.

## Declarations

### Funding

This work was subsidized by Natural Science Foundation of Ningbo Municipality (202003N4269, 2019C50069), the grants of basic public welfare projects in Zhejiang province (LGF19H020004), Zhejiang Province Medical and Health Project (2017ZD026, 2020KY273), Ningbo Health Branding Subject Fund (PPXK2018-01).

### Ethics approval and consent to participant

The Ethics Committee of the affiliated Lihuili Hospital of Ningbo University, Lihuili hospital approved this study (approval no KY2020PJ141).

### Consent for publication

Not applicable

## Competing Interests

The authors declare no competing interests.

## Data availability statement and materials

All data are fully available upon reasonable request.

## Authors' contributions

Z.F.L. and G.F.S. were responsible for study design, R.J.Z. and N.L. performed data acquisition, Z.F.L. and G.F.S. wrote the manuscript.

## Acknowledgements

Special thanks to all participants who took part in this study.

## References

1. Duffy, M. J., and O'Byrne, K. Tissue and blood biomarkers in lung cancer: A review. *Adv Clin Chem* 86, 1-21, 2018
2. Chapman, A. M., Sun, K. Y., Ruestow, P., Cowan, D. M., and Madl, A. K. Lung cancer mutation profile of EGFR, ALK, and KRAS: Meta-analysis and comparison of never and ever smokers. *Lung Cancer* 102, 122-134, 2016
3. Hong, Q. Y., Wu, G. M., Qian, G. S., et al. Prevention and management of lung cancer in China. *Cancer* 121, 3080-3088, 2015
4. Pakzad, R., Mohammadian-Hafshejani, A., Ghoncheh, M., Pakzad, I., and Salehiniya, H. The incidence and mortality of lung cancer and their relationship to development in Asia. *Transl Lung Cancer Res* 4, 763-774, 2015
5. Siegel, R. L., Miller, K. D., and Jemal, A. Cancer statistics, 2018. *CA Cancer J Clin* 68, 7-30, 2018
6. Akhtar N, Bansal JG. Risk factors of Lung Cancer in nonsmoker. *Curr Probl Cancer*. 2017 Sep-Oct;41(5):328-39
7. Tammemägi MC, Katki HA, Hocking WG, et al. Selection criteria for lung-cancer screening. *N Engl J Med*. 2013 Feb 21;368(8):728-36
8. Zuber V, Marconett CN, Shi J, et al. Pleiotropic Analysis of Lung Cancer and Blood Triglycerides. *J Natl Cancer Inst*. 2016 Aug 26;108(12): djw167
9. Duffy MJ, O'Byrne K. Tissue and Blood Biomarkers in Lung Cancer: A Review. *Adv Clin Chem*. 2018, 86:1-21

10. Argirion I, Weinstein SJ, Männistö S, Albanes D, Mondul AM. Serum Insulin, Glucose, Indices of Insulin Resistance, and Risk of Lung Cancer. *Cancer Epidemiol Biomarkers Prev.* 2017 Oct;26(10):1519-24
11. Wei L, Champman S, Li X, et al. Beliefs about medicines and non-adherence in patients with stroke, diabetes mellitus and rheumatoid arthritis: a cross-sectional study in China. *BMJ Open.* 2017 Oct 5;7(10):e017293
12. Duffy MJ, O'Byrne K. Tissue and Blood Biomarkers in Lung Cancer: A Review. *Adv Clin Chem.* 2018;86:1-21. doi: 10.1016/bs.acc.2018.05.001. Epub 2018 May 30.
13. Migliore M, Fornito M, Palazzolo M, Criscione A, Gangemi M, Borrata F, Vigneri P, Nardini M, Dunning J. Ground glass opacities management in the lung cancer screening era. *Ann Transl Med.* 2018 Mar;6(5):90. doi: 10.21037/atm.2017.07.28.
14. Lococo F, Cusumano G, Cardillo G, SICT PNR-Working Group. It's Unnecessary to Perform N1-N2 Sampling/Dissection in Predominantly-GGO cStage-I Lung Cancer? *Ann Thorac Surg.* 2021 Apr;111(4):1405-1406. doi: 10.1016/j.athoracsur.2020.05.168. Epub 2020 Aug 3.
15. Wong BS, Chiu LY, Tu DG, Sheu GT, Chan TT. Anticancer Effects of Antihypertensive L-Type Calcium Channel Blockers on Chemoresistant Lung Cancer Cells via Autophagy and Apoptosis. *Cancer Manag Res.* 2020 Mar 13;12: 1913-27
16. Yang P, Deng W, Han Y, et al. Analysis of the correlation among hypertension, the intake of  $\beta$ -blockers, and overall survival outcome in patients undergoing chemoradiotherapy with inoperable stage III non-small cell lung cancer. *Am J Cancer Res.* 2017 Apr 1;7(4):946-54.
17. Grafetstätter M, Hüsing A, González Maldonado S, et al. Plasma Fibrinogen and sP-Selectin are Associated with the Risk of Lung Cancer in a Prospective Study. *Cancer Epidemiol Biomarkers Prev.* 2019 Jul;28(7):1221-27
18. Kuang M, Peng Y, Tao X, Zhou Z, Mao H, Zhuge L, Sun Y, Zhang H. FGB and FGG derived from plasma exosomes as potential biomarkers to distinguish benign from malignant pulmonary nodules. *Clin Exp Med.* 2019 Nov;19(4):557-64
19. Dovell F, Boffetta P. Serum uric acid and cancer mortality and incidence: a systematic review and meta-analysis. *Eur J Cancer Prev.* 2018 Jul;27(4):399-405
20. Xie Y, Xu P, Liu K, et al. Hyperuricemia and gout are associated with cancer incidence and mortality: A meta-analysis based on cohort studies. *J Cell Physiol.* 2019 Aug;234(8):14364-76
21. Lin, X. J., Lu, L., Liu, L. L., et al. (2017). Blood lipids profile and Lung cancer risk in a meta-analysis of prospective cohort studies. *Journal of Clinical Lipidology*, 11(4), 1073-81
22. Şahin F, Aslan AF. Relationship between Inflammatory and Biological Markers and Lung Cancer. *J Clin Med.* 2018 Jun 25;7(7):160
23. Lyu Z, Li N, Wang G, et al. Independent and joint associations of blood lipids and lipoproteins with lung cancer risk in Chinese males: A prospective cohort study. *Int J Cancer.* 2019 Jun 15;144(12):2972-84

24. Fang R, Yang Y, Han H, et al. Analysis of risk factors for stage I lung adenocarcinoma using low-dose high-resolution computed tomography. *Oncol Lett.* 2018 Aug,16(2):2483-89
25. Migliore M, Fornito M, Palazzolo M, et al. Ground glass opacities management in the lung cancer screening era. *Ann Transl Med.* 2018 Mar,6(5):90
26. Tu SJ, Wang CW, Pan KT, Wu YC, Wu CT. Localized thin-section CT with radiomics feature extraction and machine learning to classify early-detected pulmonary nodules from lung cancer screening. *Phys Med Biol.* 2018 Mar 14,63(6):065005
27. Qiu ZX, Cheng Y, Liu D, et
28. al. Clinical, pathological, and radiological characteristics of solitary ground-glass opacity lung nodules on high-resolution computed tomography. *Ther Clin Risk Manag.* 2016 Sep 20,12:1445-53
29. Callister ME, Baldwin DR, Akram AR, et al. British Thoracic Society guidelines for the investigation and management of pulmonary nodules. *Thorax* 2015,70(Suppl 2):ii1-54,
30. Bueno J, Landeras L, Chung JH. Updated Fleischner Society Guidelines for Managing Incidental Pulmonary Nodules: Common Questions and Challenging Scenarios. *Radiographics.* 2018 Sep-Oct,38(5):1337-50
31. Horeweg N, van Rosmalen J, Heuvelmans MA, et al. Lung cancer probability in patients with CT-detected pulmonary nodules: a prespecified analysis of data from the NELSON trial of low-dose CT screening. *Lancet Oncol* 2014, 15: 1332–41.
32. Chen H, Liu CT, Hong CQ, Chu LY, Huang XY, Wei LF, Lin YW, Tian LR, Peng YH, Xu YW. Nomogram based on nutritional and inflammatory indicators for survival prediction of small cell carcinoma of the esophagus. *Nutrition.* 2021 Apr,84:111086. doi: 10.1016/j.nut.2020.111086. Epub 2020 Nov 21.
33. Luo Y, Mei D, Gong J, Zuo M, Guo X. Multiparametric MRI-Based Radiomics Nomogram for Predicting Lymphovascular Space Invasion in Endometrial Carcinoma. *J Magn Reson Imaging.* 2020 Oct,52(4):1257-1262. doi: 10.1002/jmri.27142. Epub 2020 Apr 21.

## Tables

Table 1. Baseline characteristics

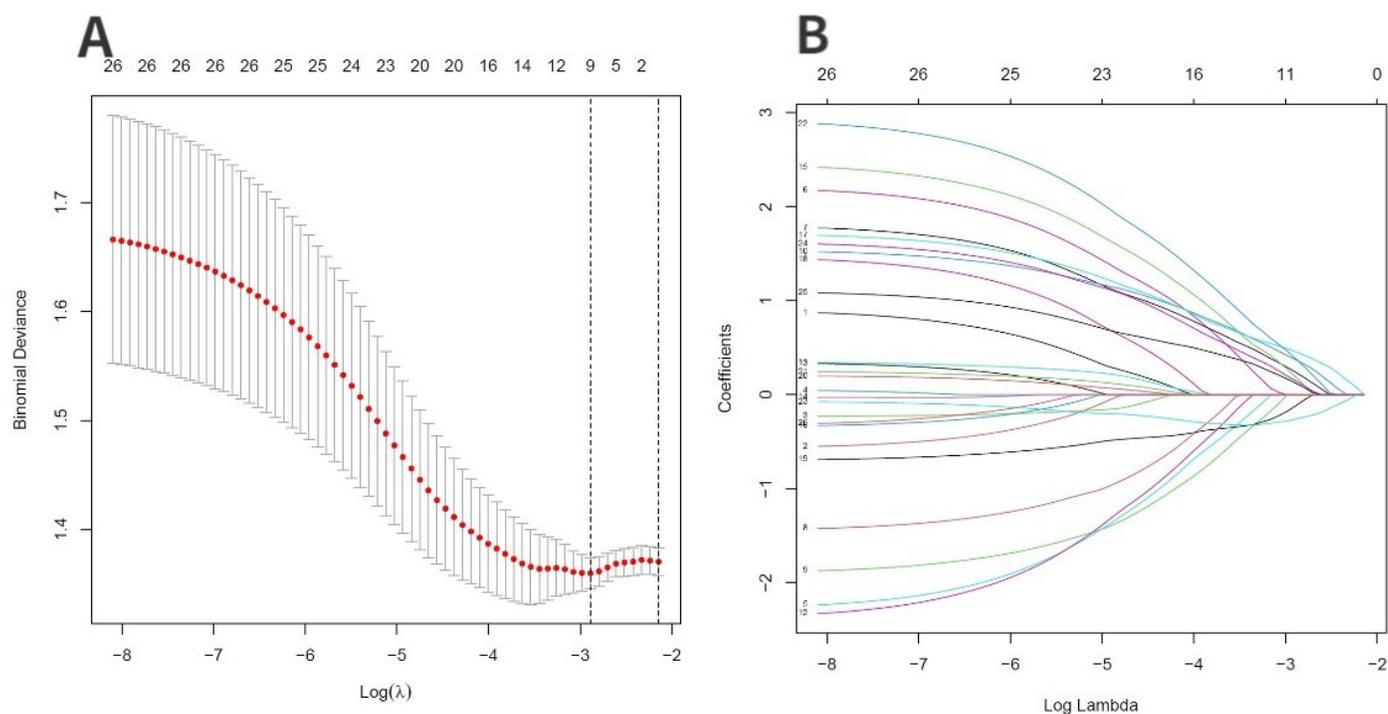
Variables	design cohort (178 Patients)		external cohort (124 patients)	
	NO.	%	NO	%
Sex				
Male/female	108/70	60.67/39.33	59/65	47.5/52.42
Age(years)				
<55/≥55	46/132	25.84/74.16	35/89	28.23/71.77
Education(years)				
Primary (0–6)	98	55.06	75	60.48
higher (>6)	80	44.94	49	39.52
BMI(Kg/m <sup>2</sup> )				
Normal (<25)	134	75.28	97	78.23
overweight or obesity (≥25)	44	24.72	27	21.77
Smoking history				
No/Yes	123/55	69.10/30.90	107/17	86.29/13.71
Drinking history				
No/Yes	146/32	82.02/17.98	112/12	90.32/9.68
Hypertension statue				
No/Yes	114/64	64.04/35.96	87/37	70.16/29.84
DM statue				
No/Yes	164/14	94.13/7.87	116/8	93.55/6.45
INR (0.89-1.12)				
Normal/Abnormal	170/8	95.51/4.49	120/4	96.77/3.23
Plasma fibrinogen level (2.38-4.98g/L)				
Normal/Abnormal	136/42	76.40/23.60	87/37	70.16/28.84
Plasma albumin (40-55g/L)				
Normal/Abnormal	80/98	44.94/55.06	71/53	57.26/42.74

GGT (10-60U/L)					
Normal/Abnormal	152/26	85.39/14.61	112/12	90.32/9.68	
CR (57-111µmol/L)					
Normal/Abnormal	115/63	64.61/35.39	72/52	58.06/41.94	
BUN 3.6-9.5mmol/L					
Normal/Abnormal	166/12	93.26/6.74	110/14	88.71/11.29	
SUA 208-428µmol/L					
Normal/Abnormal	142/36	79.78/20.22	104/20	83.87/16.13	
Blood glucose (3.9-6.1mmol/L)					
Normal/Abnormal	153/25	85.96/14.04	107/17	86.29/13.71	
HDL (1.03-1.55mmol/L)					
Normal/Abnormal	103/75	57.87/42.13	74/50	59.68/40.32	
LDL (<3.37mmol/L)					
Normal/Abnormal	157/21	88.20/11.80	109/15	87.90/12.10	
TG (<1.7mmol/L)					
Normal/Abnormal	140/38	78.65/21.35	91/33	73.39/26.61	
Location					
left lower lobe	35	19.66	24	19.35	
other lobes	88	49.44	58	46.77	
right upper lobe	55	30.90	42	33.87	
Border clear					
No/Yes	150/28	84.27/15.73	99/25	79.84/20.16	
Spicule sign					
No/Yes	156/22	87.64/12.36	110/14	88.71/11.29	
Density					
low/intermediate/high	62/22/94	34.83/12.36/52.81	38/62/24	30.65/50/19.35	
GGO					
No/Yes	116/62	65.17/34.83	72/52	58.06/41.94	

Size				
≤8mm/>8mm	32/146	17.98/82.02	23/101	18.55/81.45
Vessels pass through				
No/Yes	150/28	84.27/15.73	115/9	92.74/7.26

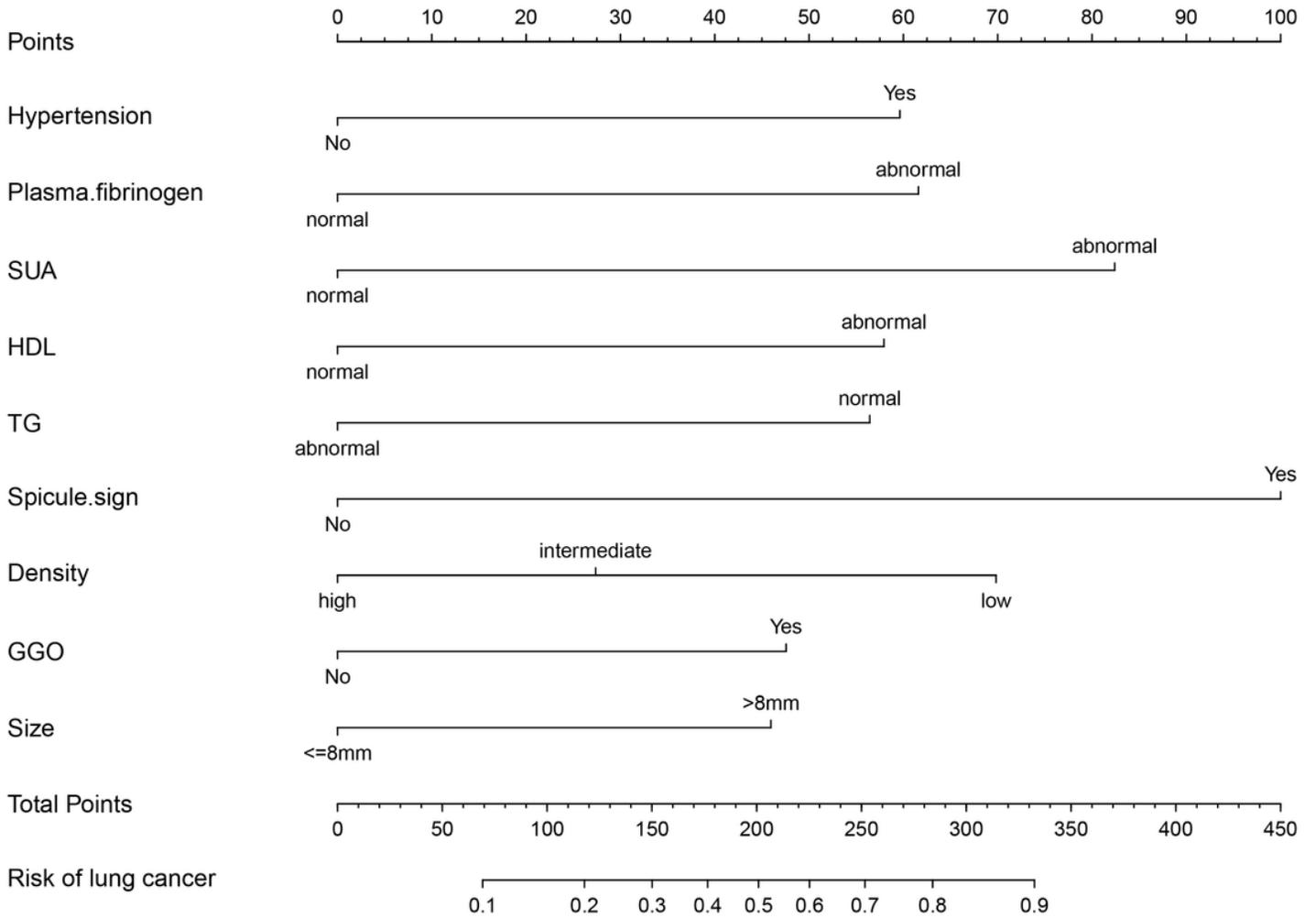
Note: The references range of different variables come from the standards of the Lihuili hospital.

## Figures



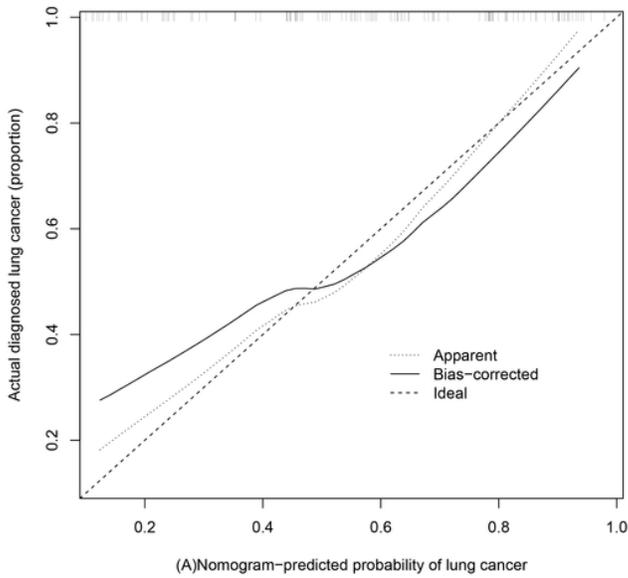
**Figure 1**

LASSO binary logistic regression model-based clinicopathological feature selection. Notes: (A) Five-fold cross-validation was used for optimal parameter (lambda) selection in the LASSO model via minimum criteria, with a partial likelihood deviance (binomial deviance) curve being plotted against log(lambda). Optimal values were marked with dashed vertical lines at optimal values using minimum criteria with the 1-SE criteria. (B) LASSO coefficient profiles for 26 potential features were generated, with coefficient profile plots against the log(lambda) sequence being generated. Five-fold validation was used to draw vertical lines at selected values, with optimal lambda resulting yielding seven total features with non-zero coefficient values. Abbreviations: SE, standard error.

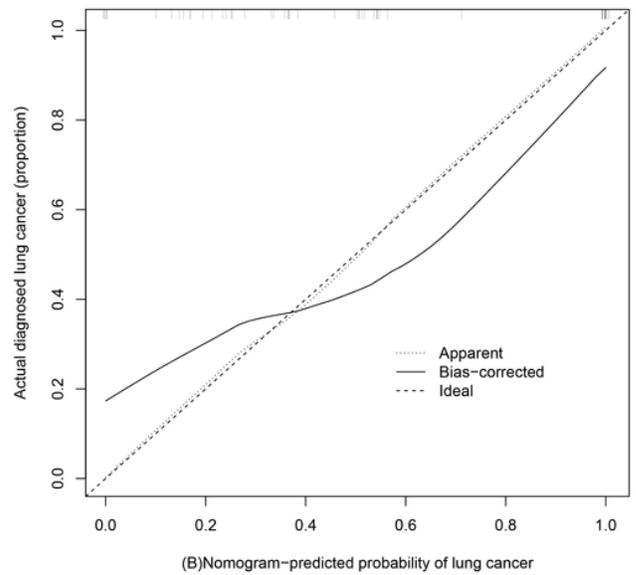


**Figure 2**

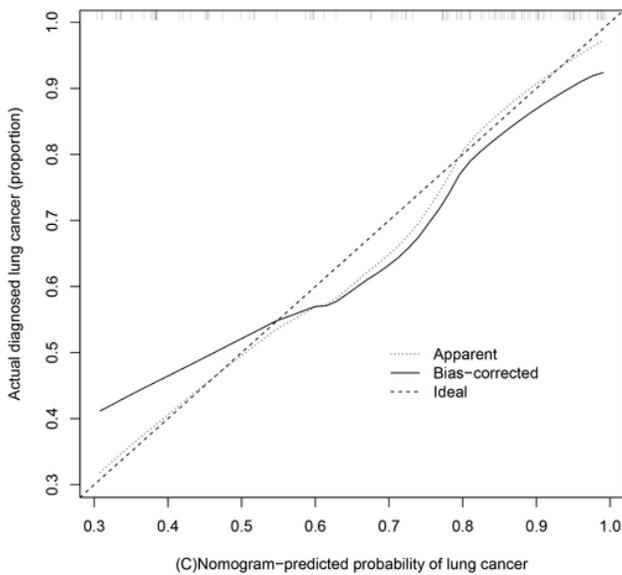
Lung cancer risk nomogram. Note: An initial training cohort was used to develop this nomogram, which incorporated SUA, hypertension status, HDL, TG, plasma fibrinogen, pulmonary nodule size, GGO status, density, and spicule sign.



**A**



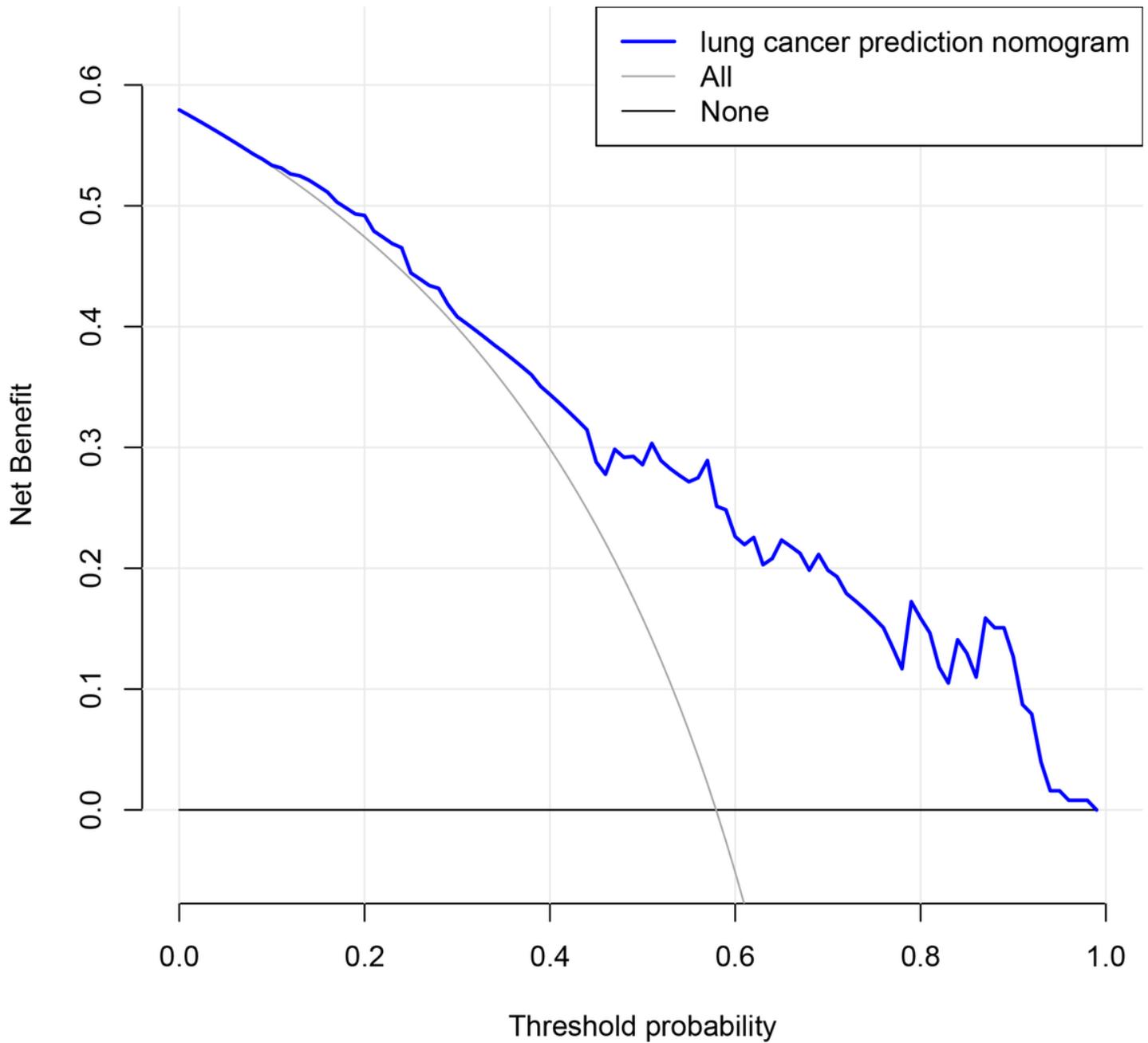
**B**



**C**

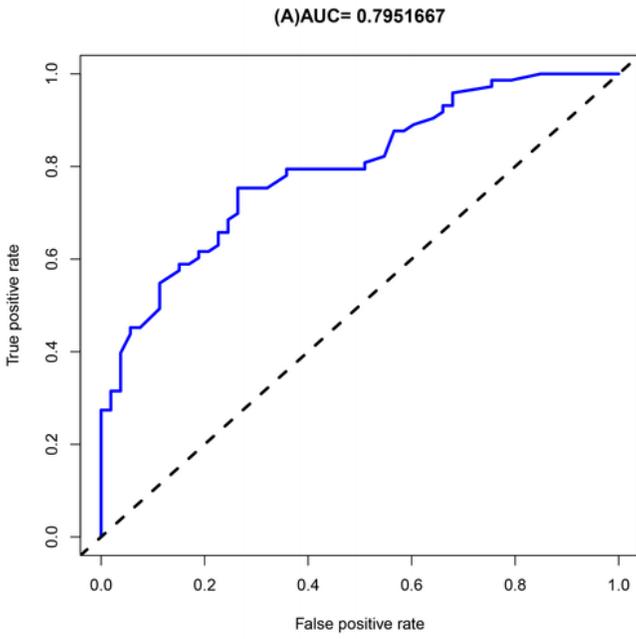
### Figure 3

Calibration curves for lung cancer nomogram predictions of training cohort (A), internal validation cohort (B) and external cohort (C), respectively. Note: Predicted risk of lung cancer and actual lung cancer diagnoses are shown on the x-axis and y-axis, respectively, with the dotted line corresponding to a diagnostic model with perfect predictive accuracy and the solid line corresponding to actual nomogram performance. The closer these lines are to one another, the better the predictive performance of this nomogram.

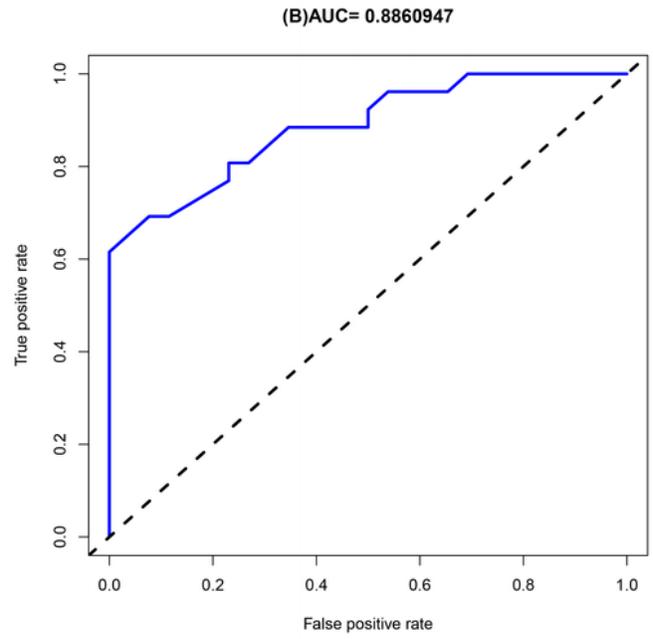


**Figure 4**

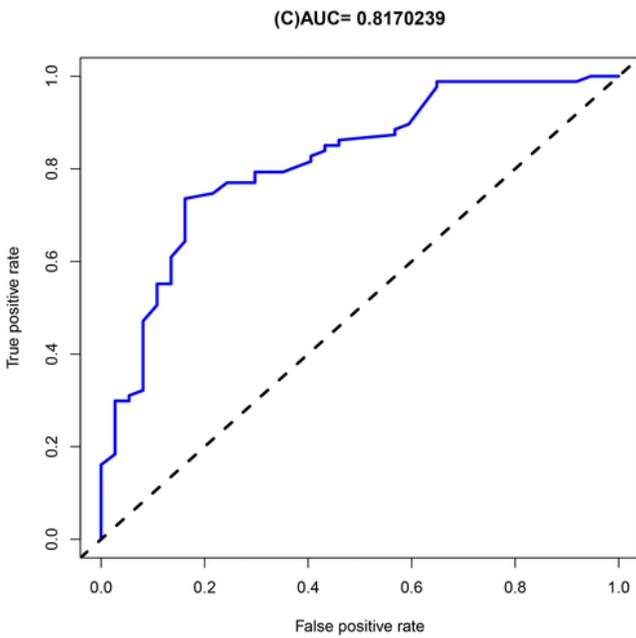
Decision curve analysis. Notes: Net benefit is shown on the y-axis, with the blue line corresponding to the lung cancer risk nomogram. The thin and thick lines respectively correspond to the assumptions that all patients or no patients have lung cancer, with the decision curve demonstrating that if the threshold probability of a patient and a doctor is >9 and <98%, respectively, then the use of this nomogram to predict the risk of lung cancer risk is more beneficial than a treat-all or treat-none interventional scheme for these patients.



A



B



C

## Figure 5

Receiver operating characteristic curve analysis for training cohort (A), internal validation cohort (B) and external cohort (C), respectively.