A Multi-Parametric Prognostic Model Based on Clinical Features and Serological Markers Predicts Overall Survival in Non-Small Cell Lung Cancer Patients with Chronic Hepatitis B Viral Infection

Shulin Chen
Sun Yat-sen University Cancer Center
Hanqing Huang
Maoming People's Hospital
Yijun Liu
Sun Yat-sen University Cancer Center
Changchun Lai
Maoming People's Hospital
Songguo Peng
Sun Yat-sen University Cancer Center
Lei Zhou
Maoming People's Hospital
Hao Chen
Sun Yat-sen University Cancer Center
Yiwei Xu
Shantou University Medical College Cancer Hospital
Xia He (hexia@sysucc.org.cn)
Sun Yat-sen University Cancer Center https://orcid.org/0000-0003-4319-4995

Research article

Keywords: Non-small cell lung cancer, Hepatitis B virus, Lasso regression, Model, Prognostic

Posted Date: September 4th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-69252/v1

Version of Record: A version of this preprint was published at Cancer Cell International on November 19th, 2020. See the published version at https://doi.org/10.1186/s12935-020-01635-8.

Abstract

features and serological markers to estimate overall survival (OS) in non-small cell lung cancer (NSCLC) patients with chronic hepatitis B viral (HBV) infection.

Methods: The prognostic model was generated by using Lasso regression in training cohort. The incremental predictive value of the model to traditional TNM staging and clinical treatment for individualized survival was evaluated by concordance index (C-index), time-dependent ROC (tdROC), and decision curve analysis (DCA). A model risk score nomogram for OS was built by combining TNM staging and clinical treatment. Then we stratified patients into high and low risk subgroups according to the model risk score. Difference in survival between subgroups was analyzed using Kaplan–Meier survival analysis. Furthermore, correlations between the prognostic model and TNM staging or treatment were analysed.

Results: The C-index values of the model for predicting OS were 0.769 and 0.676 in the training and validation cohorts, respectively, which were higher than that of TNM staging, and treatment, the tdROC curve and DCA also showed the model had good predictive accuracy and discriminatory power than TNM staging and treatment. And the nomogram shown some clinical net benefit. According to the model risk score, we divided the patients into low risk and high risk subgroups. The differences of OS rates were significant in the subgroups. Furthermore, the model was positive correlation with TNM staging and treatment.

Conclusions: The proposed prognostic model showed favorable performance than traditional TNM staging and clinical treatment for estimating OS in NSCLC (HBV+) patients.

Background

At present, lung cancer is the leading cause of cancer morbidity and mortality around world(Siegel et al. 2016), and non-small-cell lung cancer (NSCLC) accounts for 75–80% of all lung malignancies(Le Pechoux 2011). The 5-year survival of NSCLC patients is generally poor owing to late diagnosis, frequent relapse and lack of effective systemic therapy(Aggarwal et al. 2016).

Hepatitis B virus (HBV) is one of the most prevalent and most serious viral hepatitis, and there is a high prevalence of HBV in China(Cui et al. 2013). Therefore, it is reasonable to speculate that chronic HBV infection may be an important comorbidity in patients with NSCLC in China. Previous studies found that HBV has been associated with some extra-hepatic cancers(Engels et al. 2010; Liu et al. 2014; Song et al. 2019), and some cancers such as diffuse large B-cell lymphoma(Wang et al. 2008) and multiple myeloma(Teng et al. 2011) patients with HBV infection have poor survival outcomes compared to the non-infected patients. These results implied that NSCLC with HBV-infected patients should be distinguished from those uninfected patients because they have different clinical characteristics, outcomes and prognostic factors, which need to develop a distinct prognostic predictive model for the NSCLC with HBV-infected patients.

Currently, the TNM (tumor, lymph node, metastasis) is a widely used staging system for predicting the outcome of NSCLC patients(Greene et al. 2002). However, patients within the same TNM stage show different genetic, cellular, and clinicopathological characteristics, and exhibit a wide spectrum of clinical survival outcomes, indicating the need for additional prognostic factors to complement the TNM staging to better predict the outcome of the NSCLC patients(Burke 2004; Cagle 2010; Vielh et al. 2005). Therefore, much study have reported some prognostic factors that might improve the predict the survival of NSCLC(Bianconi et al. 2018; Matikas et al. 2016; Thakur et al. 2016), these findings could help identify patients that would benefit from novel therapeutic strategies or, alternatively, and whether an additional treatment needs to be carried out.

Thus, the present retrospective study was aimed to develop and validate a multi-parametric prognostic model based on clinical features and serological markers to estimate overall survival (OS) in NSCLC HBV (+) patients and assess its incremental value to the traditional staging system and clinical treatment for individual OS estimation.

Methods

Patient selection and data collection

Patients with first diagnosed NSCLC (HBV+) patients who were treated at Sun Yat-sen University Cancer Center (Guangzhou, China) between January 2008 and December 2010 were retrospectively enrolled in this study. And this study was approved by the Hospital Ethics Committee in Sun Yat-sen University Cancer Center in China. The inclusion criteria for the study are as follows: (a) pathological evidence of NSCLC; (b) patients without pathological diagnosis or with previous or concomitant malignancies; (c) hepatitis B surface antigen (HbsAg) positive; (d) without co-infected other types of hepatitis viruses; (e) complete baseline clinical information, laboratory, and follow-up data.

The following relevant clinical and serological data were collected for each enrolled patient at the time of diagnosis before any treatment: The data included age; gender, family history, body mass index (BMI), tumor size, clinical treatment, Tumor Node Metastasis stage (TNM stage)(Edge et al. 2010), white blood cell (WBC), neutrophils (N), lymphocyte (L), platelet (PLT), hepatitis B surface antigen (HbsAg), hepatitis B surface antibody (HBsAb), hepatitis B envelope antigen (HBeAg), hepatitis B envelope antibody(HBeAb), hepatitis B core antibody (HBcAb), hepatitis B core antigen (HBcAb), albumin (ALB), alkaline phosphatase (ALP), apolipoprotein AI (APOA), apolipoprotein B (APOB), C-reactive protein (CRP), lactic dehydrogenase (LDH), glutamyl transpeptidase (GGT), total bilirubin (TBIL), and direct bilirubin (DBIL). NLR was the ratio of neutrophil to lymphocyte ratio(Diem et al. 2017); PLR was the ratio of platelet to lymphocyte(Diem et al. 2017); SLR was the ratio of AST to ALT(Lee et al. 2017); ABR was the ratio of APOA to APOB(Chang et al. 2007); CAR was the ratio of C-reactive protein to albumin ratio(Deng et al. 2018); prognostic index (PI): score 0 for CRP 10 mg/L or less and white cell count 11×10^9/L or less, patients with only one of these abnormalities were allocated a score of 1, and if both of them were elevated were allocated a score of 2(Kasymjanova et al. 2010); the prognostic nutritional index (PNI) was calculated according to the

following formula: Alb (g/L) + 5 × lymphocyte count × 10^9/L: score 0 for PNI > 45; score 1 if patients with PNI ≤ 45(He et al. 2018); Glasgow prognostic score (GPS) was classified as follows: patients with serum CRP > 10 mg/L and albumin < 35 g/L were classified as GPS 2; patients with CRP > 10 mg/L or albumin < 35 g/L were classified as GPS 1; patients with serum CRP ≤ 10 mg/mL and albumin > 35 g/L were classified as GPS 0(Shiba et al. 2017).

Patients Follow Up

The patients' survival data follow-up was obtained by means of retrieving medical records, email, and direct telecommunication, all patients were followed up until death or January 2016, if still alive. The endpoint of this study was overall survival (OS), which was defined as the time interval from diagnosis to the date of the patient’s death or censored at the date of last follow-up.

Statistical Analyses

Statistical analyses were performed using IBM SPSS Statistical software version 19.0 (IBMCorp., Chicago, IL, USA) and R version 3.6.0 (http://www.R-project.org). Categorical variables were classified based on clinical findings, and continuous variables were transformed into categorical variables based on the cut-off values of by the R package "survival"(Diboun et al. 2006) and "survminer". Differences in distributions between the training cohort and validation cohort was used to Chi-square test. We utilized the Lasso regression model to select the most useful prognostic variables in the training cohort. According to the regulation weight λ, LASSO shrinks all regression coefficients towards zero and sets the coefficients of many irrelevant features exactly to zero. The optimal values of the penalty parameter λ were determined by 10-fold cross validation with the 1 standard error of the minimum criteria (the 1-SE criteria), where the final value of λ yielded minimum cross validation error. Te retained features with nonzero coefficients were used for regression model fitting(Goeman 2010; Tibshirani 1997). Then a prognostic model based on computing for each patient through a linear combination of selected variables weighted by their respective coefficients was established. The R package "glmnet" was used to perform the Lasso regression analysis. The incremental predictive value of the prognostic model to the traditional TNM staging and clinical treatment for individualized survival was evaluated by Harrell's concordance index (C-index), time-dependent ROC (tdROC), and decision curve analysis(Vickers et al. 2008). The area under the curve (AUC) was calculated using the "survivalROC" package(Heagerty et al. 2000), C-index was computed and compared by using the "survcomp" package.(Schroder et al. 2011) A nomogram (by the package of rms in R) was developed using prognostic model risk score, TNM staging, and clinical treatment. Its performance was assessed by calibration curve in internal validation with bootstrapping (1000 bootstrap resamples)(Shim et al. 2015). For subsequent comparison, we divided patients into high- and low-risk groups basing on the optimal cut-off value of prognostic model risk score, and Kaplan-Meier method and log-rank tests were used to assess differences in OS between the predicted high- and low-risk groups. The correlation between the prognostic model and TNM staging or clinical

treatment was evaluated by Pearson's correlation coefficient(Williams 1996). Results with two-sided p values of < 0.05 were considered as statistically.

Results

Patient characteristics

A total of 201 eligible patients were analyzed: 145 cases in the training cohort and 56 cases in the validation cohort. The median follow-up was 29.0 months (interquartile range (IQR):12.0–64.0) in the training cohort and 32.5 months (IQR: 11.0–60.75) in the validation cohort. The 1-, 3-, and 5-year OS rates in the training were 75.2%, 46.9%, and 31.7%, and the 1-, 3-, and 5-year OS rates in the validation were 73.2%, 42.9%, and 26.8%.

The optimal cut-off value for each continuous variable as follows: age (40 years), BMI (22.3 kg/m^2), tumor size (4.0 cm), WBC (10.8 10^9/L), N (8.1 10^9/L), L (1.74 10^9/L), PLT (163.0 10^9/L), NLR (2.7), PLR (108.6), ALB (42.5 g/L), ALT (13.7 U/L), AST (32.2 U/L), SLR (1.5), ALP (69.6 U/L), APOA (1.2 g/L), APOB (1.0 g/L), ABR (0.8), CRP (6.2 mg/L), CAR (0.16), LDH (230.3 U/L), GGT (44.2 U/L), TBIL (15.4 umol/L), DBIL (3.0 umol/L), and PNI (48.1). The details regarding patients' clinical characteristics and laboratory serological markers for the patients were listed in Table 1. No clinical and serological parameters except ALB, PLR, HBeAg, HBeAb, and HBcAb were significantly different distribution in the training cohort and validation cohort.

Construction of the multi-parametric prognostic model based on clinical and serological markers

To select prognostic clinical and serological markers, we performed the Lasso regression model on the basis of OS in the training cohort. Figure 1A showed the change in trajectory of each independent marker was analyzed. Moreover, 10-fold cross-validation was employed for model construction, and the confidence interval under each λ is presented in Fig. 1B. The optimal value of the λ was 0.046 in this model. So, this value was selected as the final model, which included 10 predictors from the 34 markers were significant weighted prognostic factors: age, BMI, tumor size, PLT, PLR, ALT, GGT, LDH, TBIL, and APOA. The coefficients of the 10 predictors were presented in Fig. 1C. Subsequently, a multi-parametric prognostic model based on clinical and serological markers was constructed using the coefficients derived from the Lasso regression model, with a prognostic model risk score calculated based on their personalized levels of the 10 predictors, by using the following formula: The prognostic model risk score = 0.679 - (0.148 × age) - (0.193 × BMI + (0.101 × tumor size) - (0.554 × PLT) + (0.197 × PLR) - (0.199 × ALT) + (0.186 × GGT) + (1.248 × LDH) - (0.137 × TBIL) - (0.194 × APOA). In this formula, each variable level was valued as 0 or 1; a value of 0 was assigned when the marker was less than or equal to the corresponding cut-off value, and a value of 1 otherwise.

Assessment Of Performance Of Prognostic Model And Verification

The C-index was used to estimate the discrimination performance between the prognostic model and TNM staging or clinical treatment. The results were listed in Table 2. In the training cohort, the C-index for prognostic model was 0.769 (95% confidence interval (CI): 0.721–0.817), which was higher than that of the TNM staging (0.710, 95% CI: 0.661–0.758, P = 0.079), and clinical treatment (0.694, 95% CI: 0.643–0.746, P = 0.017). Compared to either the TNM staging or the clinical treatment, the prognostic model also showed a better discrimination capability with higher C-indexes in the validation cohort.

The prognostic accuracy of the prognostic model and TNM staging or clinical treatment in these cohorts was also assessed using tdROC analysis (Fig. 2). In the training cohort, tdROC analysis showed that the area under ROC curve (AUC) of prognostic model was 0.857 for 1-year survival, 0.845 for 3-year survival, and 0.879 for 5-year survival, respectively. The AUC of TNM staging was 0.787 for 1-year survival, 0.798 for 3-year survival, and 0.771 for 5-year survival, respectively. The AUC of clinical treatment was 0.771 for 1-year survival, 0.799 for 3-year survival, and 0.753 for 5-year survival, respectively. The results indicated the prognostic model had better ability to predict survival outcomes than TNM staging and clinical treatment. Similar results were observed in the validation cohort.

In addition, the decision curve analysis (Fig. 3) showed the prognostic model had a higher overall net benefit than traditional TNM staging and clinical treatment across the majority of the range of reasonable threshold probabilities in training cohort and validation cohort.

Building And Validating A Predictive Nomogram

We built a nomogram consist of prognostic model risk score, TNM staging, and clinical treatment to predict 1-, 3-, and 5-year OS in the training cohort and validation cohort (Fig. 4A). Each subtype within the variables was assigned a point. As an example, locate the patient's model risk score, draw a line straight upward to the "Points" axis to determine how many points associated with that model risk score. Repeat the process for each variable, sum the points achieved for each covariate, and locate the sum on the "Total Point" axis. Final draw a line straight down to find the patient’s probability of OS at 1-, 3-, and 5-year. The calibration plots were used to assess the agreement between the predicted and actual observation at 1-, 3-, and 5-year OS (Fig. 4B, 4C, 4D). The 45° line represented the best prediction, the solid dark red line represented the performance of the nomogram in predicting the OS probability. The two lines overlap closely, indicating that the nomogram made better estimations in the patient cohort. The calibration plots for the probability of survival at 1-, 3-, and 5-year showed a good match between the prediction by nomogram and actual observation.

Performance of the prognostic model risk score in stratifying patient risk

The optimum cut-off value generated by the R package “survminer” was − 0.12 (Fig. 5). According to the cutoff values of prognostic model risk score, we divided the patients into 2 subgroups (Table 3): low risk group (risk score ≤ -0.12), and high risk group (risk score > -0.12). In the training cohort, for the high risk group, the median OS was 15 months (interquartile range (IQR): 7.0–40.0 months). And the high risk group with survival probabilities of 59.3%, 26.7% and 11.6% for 1-, 3-, and 5-year, respectively. For the low risk group, the median OS was 63 months (IQR: 38.0–74.0 months). And the high risk group with survival probabilities of 98.1%, 76.3% and 61.0% for 1-, 3-, and 5-year, respectively. In the validation cohort, low risk group also had higher survival probabilities than high risk group at 1-, 3-, and 5-year, respectively. Then, we adopt the Kaplan-Meier survival analysis according to the stratified subgroup (Fig. 6A). The Kaplan–Meier curves showed that significant differences in survival distributions were found stratified subgroup in the training cohort. We further applied it to the validation cohort, and found similar results.

Furthermore, we performed stratified analyses of NSCLC HBV (+) patients with their respective stage I/II, and III/IV (Fig. 6B, 6C). In the training cohort, the stratification by the prognostic model risk score resulted in significant differences in Kaplan–Meier OS curves for patients in each stage group. As for the validation cohort, this stratification also resulted in significant differences in OS, except for patients in stage I/II.

The correlation between the prognostic model and TNM staging or clinical treatment

Figure 7 showed the correlations between the prognostic model and TNM staging or clinical treatment in training cohort (A) and validation cohort (B). In this plot, the blue displayed positive correlations, and the red displayed negative correlations. The color intensity and the size of the circle are proportional to the correlation coefficients. In addition, the numbers in the graph show the Pearson's correlation coefficient (PCC) between different variables. The results revealed that prognostic model was positive correlation with TNM staging (PCC: training cohort: 0.48; validation cohort: 0.42) and clinical treatment (PCC: training cohort: 0.44; validation cohort: 0.29).

Discussion

In the present study, we analyzed individual clinical features and serological markers based on the survival analysis approach. Then a multi-parametric prognostic model was generated by using the Lasso regression model for predicting the overall survival in NSCLC HBV (+) patients. Our prognostic model showed better predictive accuracy and discriminative ability than traditional TNM staging and clinical treatment. The prognostic model signature successfully stratified those patients into high-risk and low-risk subgroups with significant differences in OS.

According to the results of LASSO, the present prognostic model consisting of 10 prognostic factors: age, BMI, tumor size, PLT, PLR, ALT, GGT, LDH, TBIL, and APOA. Among the 10 prognosis-specific factors, all had been reported to be associated with overall survival in lung cancer patients(Bozkaya et al. 2019; Jin et al. 2018; Kim et al. 2018; Kim et al. 2016; Li et al. 2015; Mezquita et al. 2018; Nakagawa et al. 2016; Pilling et al. 2010; Shi et al. 2018; Wang et al. 2017; Xu et al. 2015). These suggested that our analysis

results had credible prognostic value. We next compared the predictive accuracy of the prognostic model with the traditional TNM staging and clinical treatment. The C-index of the prognostic model was higher than that of TNM staging and clinical treatment both in training cohort. TdROC curve showed that our prognostic model exhibited good accuracy in clinical outcome prediction either for 1-year (AUC = 0.857), 3-year (AUC = 0.845) and 5-year (AUC = 0.879) OS of NSCLC HBV(+) patients in training cohort, when compared with traditional TNM staging and clinical treatment. The decision curve analysis also showed the prognostic model was good performance in prognosis prediction than TNM staging and clinical treatment in training cohort. Above similar results were observed in the validation cohort.

To complement the shortcomings of current TNM staging in prognostic assessment of NSCLC HBV(+) patients, the prognostic model risk score of patients was calculated, and prediction and verification were also carried out. The results showed that the prognostic model risk score successfully classified patients into high-risk and low-risk subgroups within stages I/II and III/IV, and the high-risk patients with poor survival outcomes. So, even between patients in the same stage, the high-risk patients needed more intensified treatment. These implied that the prognostic model could reinforce the prognostic ability of TNM staging. And the improved prediction of individual outcomes would be useful for counselling patients, personalizing treatment, and scheduling patients' follow-ups. Of note, there was a significant positive correlations among the prognostic model, TNM staging, and clinical treatment, suggesting the prognostic model could be useful in predicting the outcomes of NSCLC HBV (+) and might be useful in informing treatment decisions.

Compared to previous studies(Chen et al. 2018a; Chen et al. 2018b), this study had the following advantages: 1. In order to increase prognostic accuracy, many potential prognostic factors have been assessed. The potential prognostic factors included in this study were more than previously studies. 2. We developed the prognostic model using the newly algorithm LASSO model, as a statistical method for screening variables to establish the prognostic model, which enabled to adjust for model's over fitting and avoid extreme predictions. So the predictive accuracy could be improved significantly, and this approach had been applied in many study(Moons et al. 2004; Srivastava et al. 2009; Tibshirani 1997). 3. The prognostic model was different from the previous ones because of the prognostic model did not include the TNM staging. Therefore, it can be used for patients with TNM staging was unclear. Moreover, the C-index of the prognostic model was approximately equivalent or even higher than the previously reported model. 4. In the published articles, the continuous variables need to be transformed into categorical variables based on the cut-off values for the further research. There had some limitations in choosing the cut-off values for the continuous variables, because the cut-off values were determined by the analyzed data, different analysis data have different cut-off values. In order to overcome this limitation, in this study, the continuous variables did not need to be transformed into categorical variables. So, this was convenient for other center applications.

However, some limitations in our study should be considered. First, this was a retrospective analysis, and thus the retrospective nature of this study cannot exclude all potential biases. Second, our endpoint was overall survival, and further research should also be conducted on the disease-free survival (DFS). Third,

other predictive biomarkers such as radiomics features(Huang et al. 2016), carcinoembryonic antigen (CEA)(Baek et al. 2018), cytokeratin 19 fragment (CYFRA21-1)(Baek et al. 2018), epidermal growth factor receptor (EGFR)(Romero-Ventosa et al. 2015), circulating tumor cells(Coco et al. 2017), and circulating cell-free DNA(Zhang et al. 2018) were not analyzed in the study. Finally, analysis of data from a single cancer center, and the sample sizes were small. A large-scale and multicenter validation of these results will be needed in the future. Despite the above shortcomings, the prognostic model was effective and may be useful in predicting the outcomes of NSCLC HBV (+) patients.

Conclusions

In conclusion, this study provided a multi-parametric prognostic model derived from clinical features and serological markers that showed favorable performance than traditional TNM staging and clinical treatment for individualized OS estimation, and the nomogram based on prognostic model, TNM staging, and clinical treatment can reinforce the prognostic ability of TNM staging. Therefore, the simple, precise and understandable prognostic model may serve as a potential tool for clinicians in counselling patients, personalizing treatment, and scheduling patients' follow-ups for NSCLC HBV (+) patients.

Abbreviations

AUC = the areas under ROC curve; BMI = body mass index; TNM = Tumor Node Metastasis stage; Sur = surgery; Rad = radiotherapy; Che = chemotherapy; WBC = white blood cell; NLR = neutrophil/lymphocyte ratio; PLR = platelet/lymphocyte ratio; ALB = albumin; ALT = alanine transaminase; AST = aspartate aminotransferase; SLR = AST/ALT ratio; ALP = alkaline phosphatase; APOA = apolipoprotein AI ; APOB = apolipoprotein B; ABR = APOA/APOB ratio; CRP = C-reactive protein; CAR = C-reactive protein/albumin ratio; LDH = lactic dehydrogenase; GGT = glutamyl transpeptidase; TBIL = total bilirubin; DBIL = direct bilirubin; PNI = prognostic nutritional index; PI = prognostic index; PCC = Pearson's correlation coefficient; GPS = Glasgow prognostic score; IQR = interquartile range.

Declarations

Availability of data and materials

The datasets analyzed during the current study are not publicly available due to patient privacy concerns, but are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

This study was approved by the Clinical Research Ethics Committee of the Sun Yat-sen University Cancer Center, and all patients provided written informed consent at the first visit to our center.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China (no. 31600632) and funded by the Natural Science Foundation of Guangdong Province (2018A030307079).

Authors' contributions

All authors contributed to this manuscript, including conception and design (HC, YWX, XH), acquisition of data (SLC, YJL, SGP), analysis and interpretation of data (HQH, CCL, LZ), material support (SLC), study supervision (HC, YWX, XH), and all authors read and approved the final manuscript.

Acknowledgments

We thank Sun Yat-sen University Cancer Center for providing support on research conditions in this study.

References

1. Aggarwal A, Lewison G, Idir S, et al. The State of Lung Cancer Research: A Global Analysis. J Thorac Oncol. 2016;11:1040–50.
2. Baek AR, Seo HJ, Lee JH, et al. Prognostic value of baseline carcinoembryonic antigen and cytokeratin 19 fragment levels in advanced non-small cell lung cancer. Cancer Biomark. 2018;22:55–62.
3. Bianconi F, Fravolini ML, Bello-Cerezo R, et al. Evaluation of Shape and Textural Features from CT as Prognostic Biomarkers in Non-small Cell Lung Cancer. Anticancer Res. 2018;38:2155–60.
4. Bozkaya Y, Yazici O. Prognostic significance of gamma-glutamyl transferase in patients with metastatic non-small cell lung cancer. Expert Rev Mol Diagn. 2019;19:267–72.
5. Burke HB. Outcome prediction and the future of the TNM staging system. J Natl Cancer Inst. 2004;96:1408–9.
6. Cagle PT, Lung Carcinoma Staging Problems. Surg Pathol Clin. 2010;3:61–9.
7. Chang SJ, Hou MF, Tsai SM, et al. The association between lipid profiles and breast cancer among Taiwanese women. Clin Chem Lab Med. 2007;45:1219–23.
8. Chen S, Lai Y, He Z, et al. Establishment and validation of a predictive nomogram model for non-small cell lung cancer patients with chronic hepatitis B viral infection. J Transl Med. 2018a;16:116.

9. Chen S, Li X, Lv H, et al. Prognostic Dynamic Nomogram Integrated with Inflammation-Based Factors for Non-Small Cell Lung Cancer Patients with Chronic Hepatitis B Viral Infection. Int J Biol Sci. 2018b;14:1813–21.
10. Coco S, Alama A, Vanni I, et al. Circulating Cell-Free DNA and Circulating Tumor Cells as Prognostic and Predictive Biomarkers in Advanced Non-Small Cell Lung Cancer Patients Treated with First-Line Chemotherapy. Int J Mol Sci. 2017; 18.
11. Cui Y, Jia J. Update on epidemiology of hepatitis B and C in China. J Gastroenterol Hepatol. 2013;28(Suppl 1):7–10.
12. Deng TB, Zhang J, Zhou YZ, et al. The prognostic value of C-reactive protein to albumin ratio in patients with lung cancer. Med (Baltim). 2018;97:e13505.
13. Diboun I, Wernisch L, Orengo CA, et al. Microarray analysis after RNA amplification can detect pronounced differences in gene expression using limma. BMC Genom. 2006;7:252.
14. Diem S, Schmid S, Krapf M, et al. Neutrophil-to-Lymphocyte ratio (NLR) and Platelet-to-Lymphocyte ratio (PLR) as prognostic markers in patients with non-small cell lung cancer (NSCLC) treated with nivolumab. Lung cancer. 2017;111:176–81.
15. Edge SB, Compton CC, The American Joint Committee on Cancer. the 7th edition of the AJCC cancer staging manual and the future of TNM. Ann Surg Oncol. 2010;17:1471–4.
16. Engels EA, Cho ER, Jee SH. Hepatitis B virus infection and risk of non-Hodgkin lymphoma in South Korea: a cohort study. Lancet Oncol. 2010;11:827–34.
17. Goeman JJ. L1 penalized estimation in the Cox proportional hazards model. Biom J. 2010;52:70–84.
18. Greene FL, Sobin LH. The TNM system: our language for cancer care. J Surg Oncol. 2002;80:119–20.
19. He X, Li JP, Liu XH, et al. Prognostic value of C-reactive protein/albumin ratio in predicting overall survival of Chinese cervical cancer patients overall survival: comparison among various inflammation based factors. J Cancer. 2018;9:1877–84.
20. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics. 2000;56:337–44.
21. Huang Y, Liu Z, He L, et al. Radiomics Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early-Stage (I or II) Non-Small Cell Lung Cancer. Radiology. 2016;281:947–57.
22. Jin S, Cao S, Xu S, et al. Clinical impact of pretreatment prognostic nutritional index (PNI) in small cell lung cancer patients treated with platinum-based chemotherapy. Clin Respir J. 2018;12:2433–40.
23. Kasymjanova G, MacDonald N, Agulnik JS, et al. The predictive value of pre-treatment inflammatory markers in advanced non-small-cell lung cancer. Curr Oncol. 2010;17:52–8.
24. Kim JH, Seo SW, Chung CH. What Factors Are Associated With Early Mortality in Patients Undergoing Femur Surgery for Metastatic Lung Cancer? Clin Orthop Relat Res. 2018;476:1815–22.
25. Kim SH, Lee HW, Go SI, et al. Clinical significance of the preoperative platelet count and platelet-to-lymphocyte ratio (PLT-PLR) in patients with surgically resected non-small cell lung cancer. Oncotarget. 2016;7:36198–206.

26. Le Pechoux C. Role of postoperative radiotherapy in resected non-small cell lung cancer: a reassessment based on new data. Oncologist. 2011;16:672–81.
27. Lee H, Choi YH, Sung HH, et al. De Ritis Ratio (AST/ALT) as a Significant Prognostic Factor in Patients With Upper Tract Urothelial Cancer Treated With Surgery. Clin Genitourin Cancer. 2017;15:e379–85.
28. Li N, Xu M, Cai MY, et al. Elevated serum bilirubin levels are associated with improved survival in patients with curatively resected non-small-cell lung cancer. Cancer Epidemiol. 2015;39:763–8.
29. Liu X, Li X, Jiang N, et al. Prognostic value of chronic hepatitis B virus infection in patients with nasopharyngeal carcinoma: analysis of 1301 patients from an endemic area in China. cancer. 2014;120:68–76.
30. Matikas A, Syrigos KN, Agelaki S. Circulating Biomarkers in Non-Small-Cell Lung Cancer: Current Status and Future Challenges. Clin Lung Cancer. 2016;17:507–16.
31. Mezquita L, Auclin E, Ferrara R, et al. Association of the Lung Immune Prognostic Index With Immune Checkpoint Inhibitor Outcomes in Patients With Advanced Non-Small Cell Lung Cancer. JAMA Oncol. 2018;4:351–7.
32. Moons KG, Donders AR, Steyerberg EW, et al. Penalized maximum likelihood estimation to directly adjust diagnostic and prognostic prediction models for overoptimism: a clinical example. J Clin Epidemiol. 2004;57:1262–70.
33. Nakagawa T, Toyazaki T, Chiba N, et al. Prognostic value of body mass index and change in body weight in postoperative outcomes of lung cancer surgery. Interact Cardiovasc Thorac Surg. 2016;23:560–6.
34. Pilling JE, Dusmet ME, Ladas G, et al. Prognostic factors for survival after surgical palliation of malignant pleural effusion. J Thorac Oncol. 2010;5:1544–50.
35. Romero-Ventosa EY, Blanco-Prieto S, Gonzalez-Pineiro AL, et al. Pretreatment levels of the serum biomarkers CEA, CYFRA 21 – 1, SCC and the soluble EGFR and its ligands EGF, TGF-alpha, HB-EGF in the prediction of outcome in erlotinib treated non-small-cell lung cancer patients. Springerplus. 2015;4:171.
36. Schroder MS, Culhane AC, Quackenbush J, et al. survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. Bioinformatics. 2011;27:3206–8.
37. Shi H, Huang H, Pu J, et al. Decreased pretherapy serum apolipoprotein A-I is associated with extent of metastasis and poor prognosis of non-small-cell lung cancer. Onco Targets Ther. 2018;11:6995–7003.
38. Shiba H, Horiuchi T, Sakamoto T, et al. Glasgow prognostic score predicts therapeutic outcome after hepatic resection for hepatocellular carcinoma. Oncol Lett. 2017;14:293–8.
39. Shim JH, Jun MJ, Han S, et al. Prognostic nomograms for prediction of recurrence and survival after curative liver resection for hepatocellular carcinoma. Ann Surg. 2015;261:939–46.
40. Siegel RL, Miller KD, Jemal A, Cancer statistics. 2016. CA Cancer J Clin. 2016; 66:7–30.

41. Song C, Lv J, Liu Y, et al. Associations Between Hepatitis B Virus Infection and Risk of All Cancer Types. JAMA Netw Open. 2019;2:e195718.
42. Srivastava S, Chen L Comparison between the stochastic search variable selection and the least absolute shrinkage and selection operator for genome-wide association studies of rheumatoid arthritis. BMC Proc. 2009; 3 Suppl 7:S21.
43. Teng CJ, Liu HT, Liu CY, et al. Chronic hepatitis virus infection in patients with multiple myeloma: clinical characteristics and outcomes. Clinics. 2011;66:2055–61.
44. Thakur MK, Gadgeel SM. Predictive and Prognostic Biomarkers in Non-Small Cell Lung Cancer. Semin Respir Crit Care Med. 2016;37:760–70.
45. Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med. 1997;16:385–95.
46. Vickers AJ, Cronin AM, Elkin EB, et al. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. BMC Med Inform Decis Mak. 2008;8:53.
47. Vielh P, Spano JP, Grenier J, et al. Molecular prognostic factors in resectable non-small cell lung cancer. Crit Rev Oncol Hematol. 2005;53:193–7.
48. Wang F, Xu RH, Luo HY, et al. Clinical and prognostic analysis of hepatitis B virus infection in diffuse large B-cell lymphoma. BMC Cancer. 2008;8:115.
49. Wang YQ, Zhi QJ, Wang XY, et al. Prognostic value of combined platelet, fibrinogen, neutrophil to lymphocyte ratio and platelet to lymphocyte ratio in patients with lung adenosquamous cancer. Oncol Lett. 2017;14:4331–8.
50. Williams S. Pearson's correlation coefficient. N Z Med J. 1996;109:38.
51. Xu S, Xi J, Jiang W, et al. Solid component and tumor size correlate with prognosis of stage IB lung adenocarcinoma. Ann Thorac Surg. 2015;99:961–7.
52. Zhang Y, Zheng H, Zhan Y, et al. Detection and application of circulating tumor cell and circulating tumor DNA in the non-small cell lung cancer. Am J Cancer Res. 2018;8:2377–86.

Tables

Table 1. Demographics and clinical characteristics of patients in the training and

validation cohort

Characteristic	Training cohort	Validation cohort	*P* value
	n=(145)	n=(56)	
	No. (%)	No. (%)	
Gender			0.811
Male	109 (75.2%)	43 (76.8%)	
Female	36 (24.8%)	13 (23.2%)	
Age (years)			0.431
≤40	15 (10.3%)	8 (14.3%)	
>40	130 (89.7%)	48(85.7%)	
Family history			0.898
Yes	35 (24.1%)	14 (25.0%)	
No	110 (75.9%)	42 (75.0%)	
Smoking behavior			0.819
Yes	88 (60.7%)	33 (58.9%)	
No	57 (39.3%)	23 (41.1%)	
BMI (kg/m^2)			0.630
≤22.3	80 (55.2%)	33 (58.9%)	
>22.3	65 (44.8%)	23 (41.1%)	
TNM stage[a]			0.846
I	34 (23.4%)	11 (19.6%)	
II	14 (9.7%)	5 (8.9%)	
III	50 (34.5%)	23 (41.1%)	
IV	47 (32.4%)	17 (30.4%)	
Size (cm)[b]			0.911
≤4.0	79 (54.5%)	31 (55.4%)	
>4.0	66 (45.5%)	25 (44.6%)	
Treatment			0.648
Sur	31 (21.4%)	13 (23.2%)	
Sur and Rad/Che	47 (32.4%)	18 (32.1%)	

Rad/Che	54 (37.2%)	17 (30.4%)	
Other	13 (9.0%)	8 (14.3%)	
WBC (10^9/L)			0.560
≤10.8	125 (86.2%)	50 (89.3%)	
>10.8	20 (13.8%)	6 (10.7%)	
Neutrophils (10^9/L)			
≤8.1	128 (88.3%)	51 (91.1%)	0.569
>8.1	17 (11.7%)	5 (8.9%)	
Lymphocyte (10^9/L)			
≤1.74	40 (27.6%)	17 (30.4%)	0.696
>1.74	105 (72.4%)	39 (69.6%)	
Platelet (10^9/L)			0.245
≤163.0	22 (15.2%)	5 (8.9%)	
>163.0	123 (84.8%)	51 (91.1%)	
NLR			0.521
≤2.7	90 (62.1%)	32 (57.1%)	
>2.7	55 (37.9%)	24 (42.9%)	
PLR			0.043
≤108.6	64 (44.1%)	16 (28.6%)	
>108.6	81 (55.9%)	40 (71.4%)	
HBsAb			0.107
Negative	145 (100.0%)	55 (98.2%)	
Positive	0 (0.0%)	1 (1.8%)	
HBeAg			0.026
Negative	142 (97.9%)	51 (91.1%)	
Positive	3 (2.1%)	5 (8.9%)	
HBeAb			0.016
Negative	8 (5.5%)	9 (16.1%)	

Positive	137 (94.5%)	47 (83.9%)	
HBcAb			0.022
Negative	0 (0.0%)	2 (3.6%)	
Positive	145 (100.0%)	54 (96.4%)	
ALB (g/L)			0.035
≤42.5	91 (62.8%)	26 (46.4%)	
>42.5	54 (37.2%)	30 (53.6%)	
ALT (U/L)			0.260
≤13.7	26 (17.9%)	14 (25.0%)	
>13.7	119 (82.1%)	42 (75.0%)	
AST (U/L)			0.370
≤32.2	124 (85.5%)	45 (80.4%)	
>32.2	21 (14.5%)	11 (19.6%)	
SLR			0.987
≤1.5	127 (87.6%)	49 (87.5%)	
>1.5	18 (12.4%)	7 (12.5%)	
ALP (U/L)			0.931
≤69.6	56 (38.6%)	22 (39.3%)	
>69.6	89 (61.4%)	34 (60.7%)	
APOA (g/L)			0.425
≤1.2	79 (54.5%)	27 (48.2%)	
>1.2	66 (45.5%)	29 (51.8%)	
APOB (g/L)			0.315
≤1.0	101 (69.7%)	43 (76.9%)	
>1.0	44 (30.3%)	13 (23.2%)	
ABR			0.057
≤0.8	14 (9.7%)	1 (1.8%)	
>0.8	131 (90.3%)	55 (98.2%)	
CRP (mg/L)			0.443

≤6.2	82 (56.6%)	35 (62.5%)	
>6.2	63 (43.6%)	21 (37.5%)	
CAR			0.278
≤0.16	81 (17.9%)	36 (64.3%)	
>0.16	64 (44.1%)	20 (35.7%)	
LDH (U/L)			0.755
≤230.3	119 (82.1%)	47 (83.9%)	
>230.3	26 (17.9%)	9 (16.1%)	
GGT (U/L)			0.731
≤44.2	116 (80.0%)	46 (82.1%)	
>44.2	29 (20.0%)	10 (17.9%)	
TBIL (umol/L)			0.652
≤15.4	115 (79.3%)	46 (82.1%)	
>15.4	30 (20.7%)	10 (17.9%)	
DBIL (umol/L)			0.813
≤3.0	57 (39.3%)	21 (37.5%)	
>3.0	88 (60.7%)	35 (62.5%)	
PNI			0.084
≤48.1	34 (23.4%)	7 (12.5%)	
>48.1	111 (76.6%)	49 (87.5%)	
PI			0.521
0	92 (63.4%)	40 (71.4%)	
1	39 (26.9%)	11 (19.7%)	
2	14 (9.7%)	5 (8.9%)	
GPS			0.456
0	93 (64.1%)	41 (73.2%)	
1	46 (31.7%)	13 (23.2%)	
2	6 (4.1%)	2 (3.6%)	

a: TNM stage was classified according to the AJCC 7th TNM staging system;

b: The tumor maximum diameter;

Abbreviations: BMI: body mass index; TNM: Tumor Node Metastasis stage; Sur: surgery; Rad: radiotherapy; Che: chemotherapy; WBC: white blood cell; NLR: neutrophil/lymphocyte ratio; PLR : platelet/lymphocyte ratio; ALB: albumin; ALT: alanine transaminase; AST: aspartate aminotransferase; SLR: AST/ALT ratio; ALP: alkaline phosphatase; APOA: apolipoprotein AI ; APOB: apolipoprotein B; ABR: APOA/APOB ratio; CRP: C-reactive protein; CAR: C-reactive protein/albumin ratio; LDH: lactic dehydrogenase; GGT: glutamyl transpeptidase; TBIL: total bilirubin; DBIL: direct bilirubin; PNI: prognostic nutritional index; PI: prognostic index; GPS: Glasgow prognostic score.

Table 2. The C-index of our model, TNM staging and Treatment for prediction of OS in the training cohort and validation cohort

Factors	C-index (95% CI)	*P*
For training cohort		
Our model	0.769 (0.721~0.817)	
TNM staging	0.710 (0.661~0.758)	
Treatment	0.694 (0.643~0.746)	
Our model + TNM staging	0.784 (0.739~0.830)	
Our model vs TNM staging		0.079
Our model vs Treatment		0.017
Our model vs Our model + TNM staging		0.218
For validation cohort		
Our model	0.676 (0.556~0.796)	
TNM staging	0.654 (0.552~0.755)	
Treatment	0.647 (0.517~0.777)	
Our model + TNM staging	0.712 (0.614~0.809)	
Our model vs TNM staging		0.761
Our model vs Treatment		0.754
Our model vs Our model + TNM staging		0.205

C-index = concordance index; CI = confidence interval; P values are calculated based on normal approximation using function rcorrp.cens in Hmisc package.

Table 3. OS and OS rate in high-risk and low-risk groups according to our model risk score in the training and validation cohort

Parameter	Training cohort			Validation cohort		
	High-Risk Group	Low-Risk Group	Total	High-Risk Group	Low -Risk Group	Total
No. of patients	86	59	145	28	28	56
OS						
Median	15.0	63.0		12.5	59.0	
IQR	7.0 - 40.0	38.0 -74.0		7.25 - 21.50	33.50 - 72.25	
No. of OS						
At 1 year	51 (59.3%)	58 (98.1%)	109 (69.7%)	16 (57.1%)	25 (89.3%)	41 (73.2%)
At 3 year	23 (26.7%)	45 (76.3%)	68 (46.9%)	4 (14.3%)	20 (71.4%)	24 (42.9%)
At 5 year	10 (11.6%)	36 (61.0%)	46 (31.7%)	1 (3.6%)	14 (50.0%)	15 (26.8%)

Abbreviations: OS: overall survival; IQR: interquartile range.

Figures

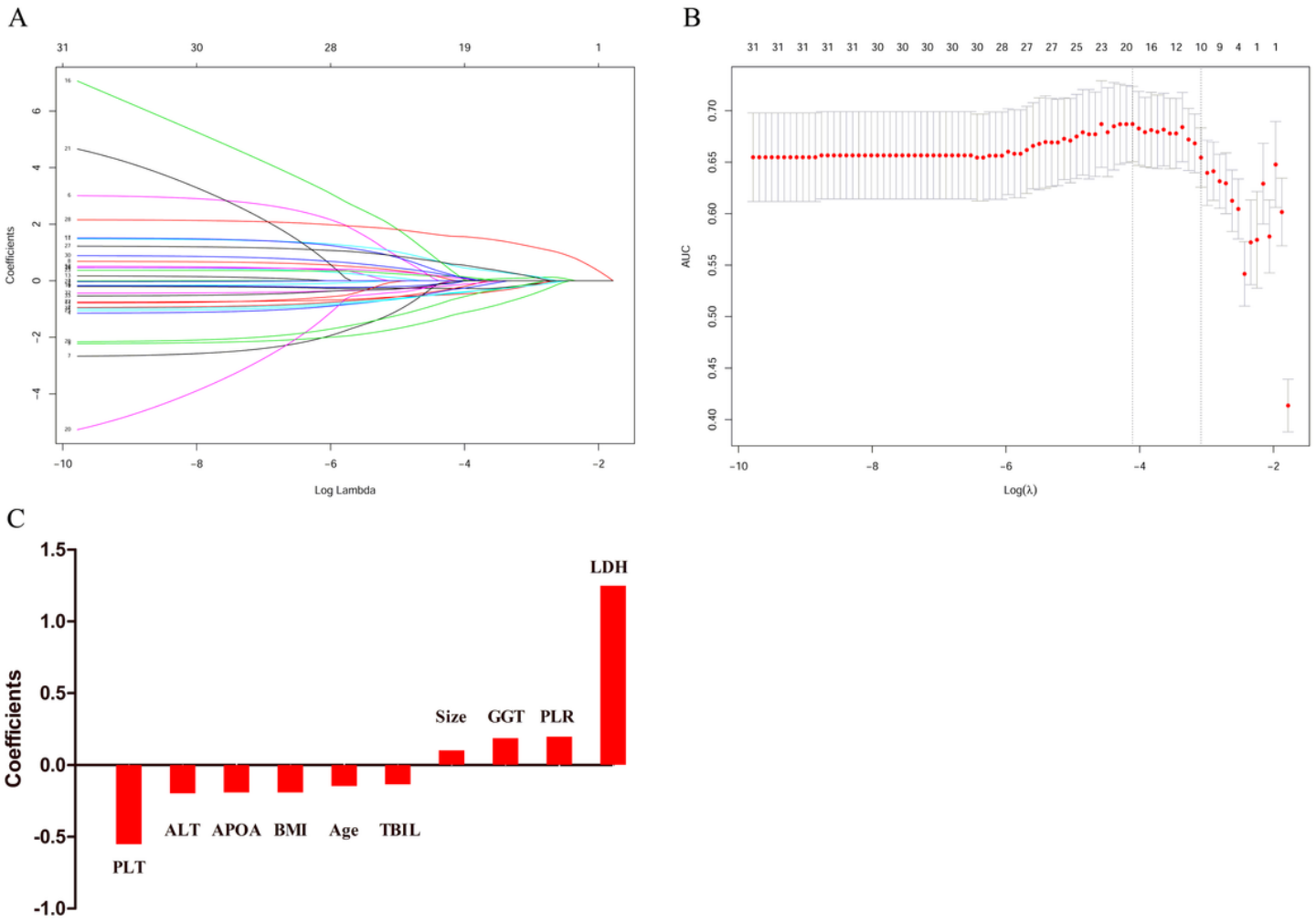

Figure 1

Potential predictors selection using LASSO regression model. A: The changing trajectory of each predictor. The horizontal axis represents the log value of the each predictor λ, and the vertical axis represents the coefficient of the independent predictor; B: Tuning the penalty parameter in LASSO using 10-fold cross validation and 1 standard error of the minimum criteria; C: Histogram showed the role of each predictor that contributed to the developed prognostic model. The predictors that contribute to the prognostic model were plotted on the x-axis, with their coefficients in the LASSO regression model plotted on the y-axis.

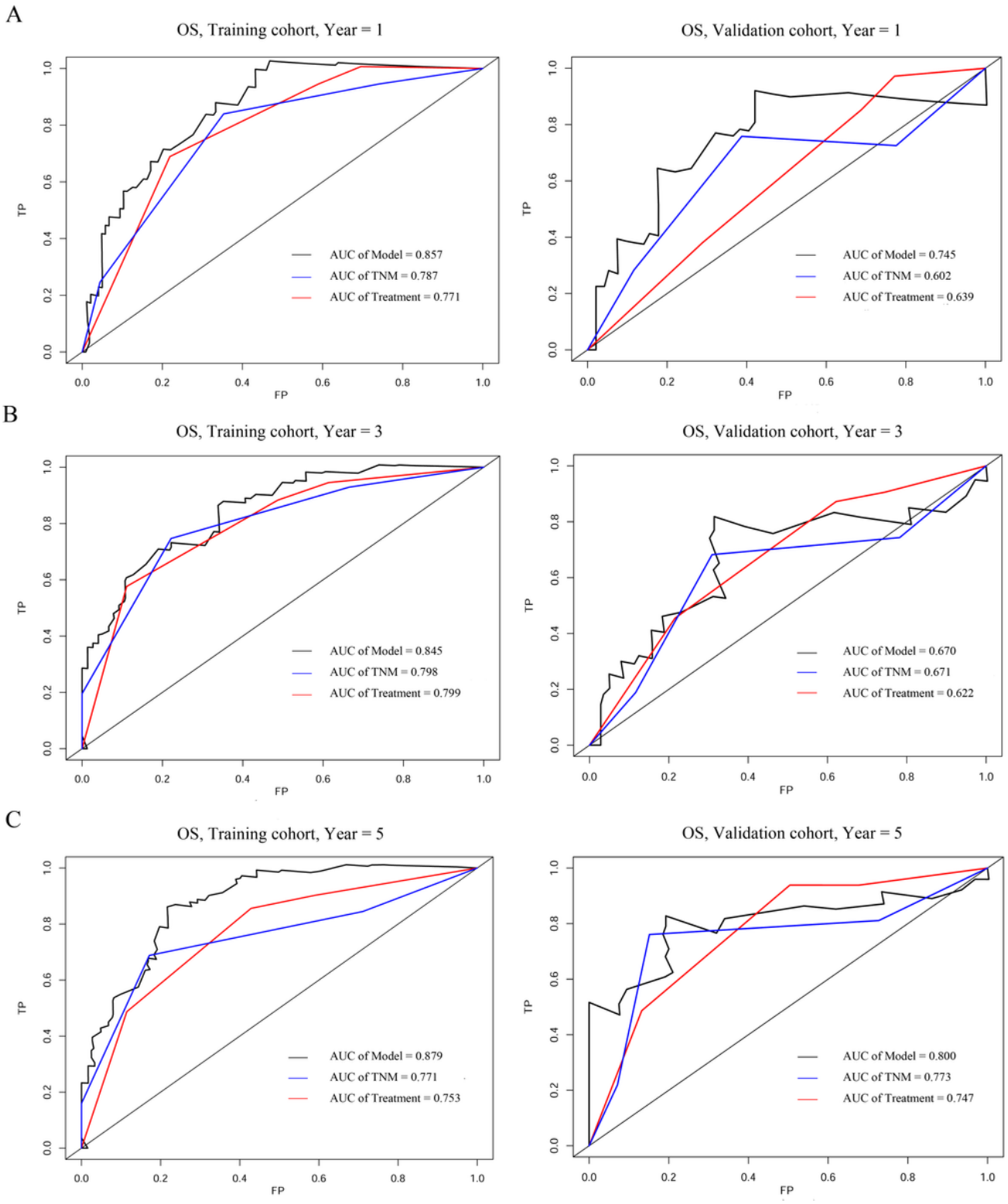

Figure 2

Comparison of predictive accuracy between prognostic model, TNM staging, and clinical treatment using time dependent ROC curves at 1-, 3-, 5- year (A, B, C) in training cohort (left) and validation cohort (right).

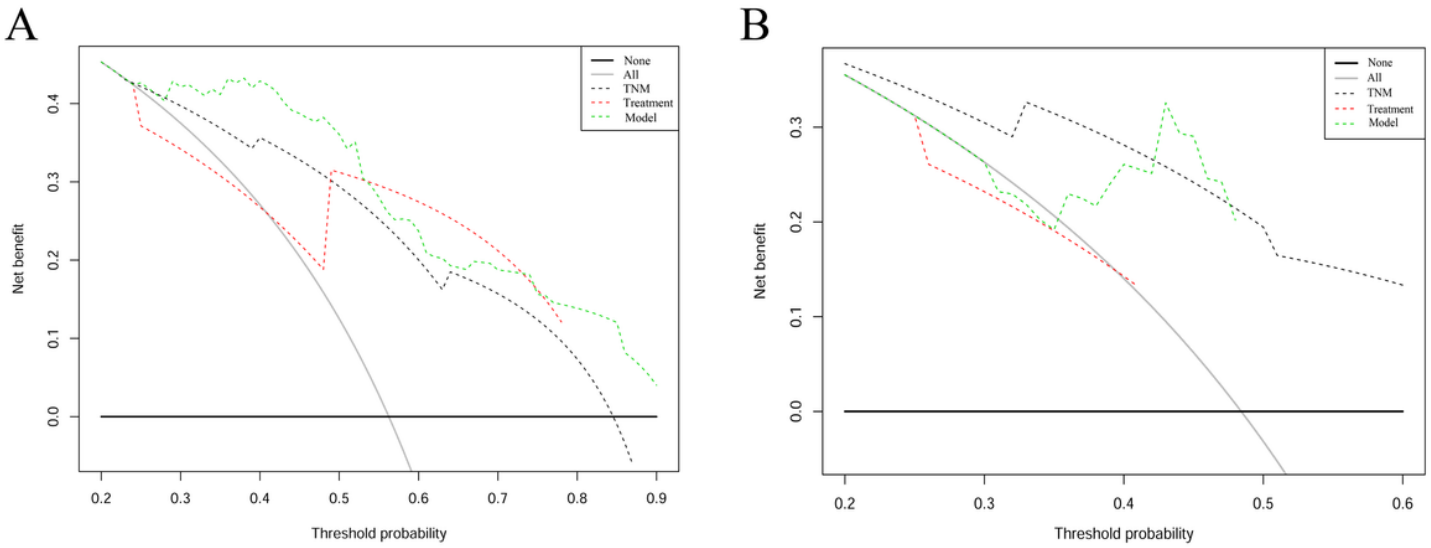

Figure 3

Decision curve analysis for each model in training cohort (A) and validation cohort (B). The thick grey line is the net benefit for a strategy of treating all men; the thick black line is the net benefit of treating no men. The y-axis indicated the net benefit, which was calculated by summing the benefits (true positive results) and subtracting the harms (false positive results), weighting the latter by a factor related to the relative harm of an undetected cancer compared with the harm of unnecessary treatment.

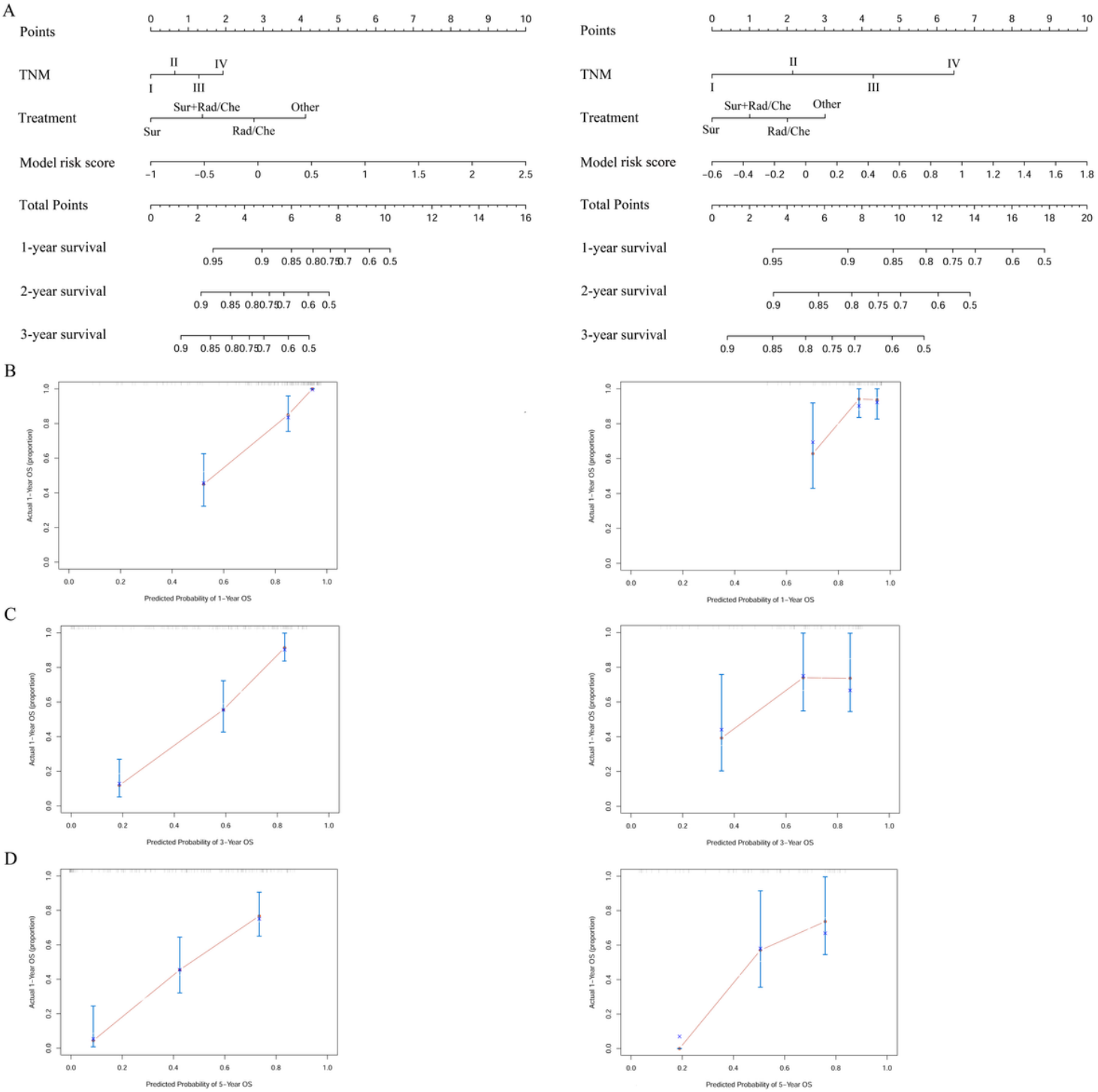

Figure 4

The nomograms (A) was used to estimate OS for NSCLC (HBV+) patients, along with the calibration plot (B, C, D) for the nomograms at 1-, 3-, 5- year in training cohort (left) and validation cohort (right).The nomograms (A) was used to estimate OS for NSCLC (HBV+) patients, along with the calibration plot (B, C, D) for the nomograms at 1-, 3-, 5- year in training cohort (left) and validation cohort (right).

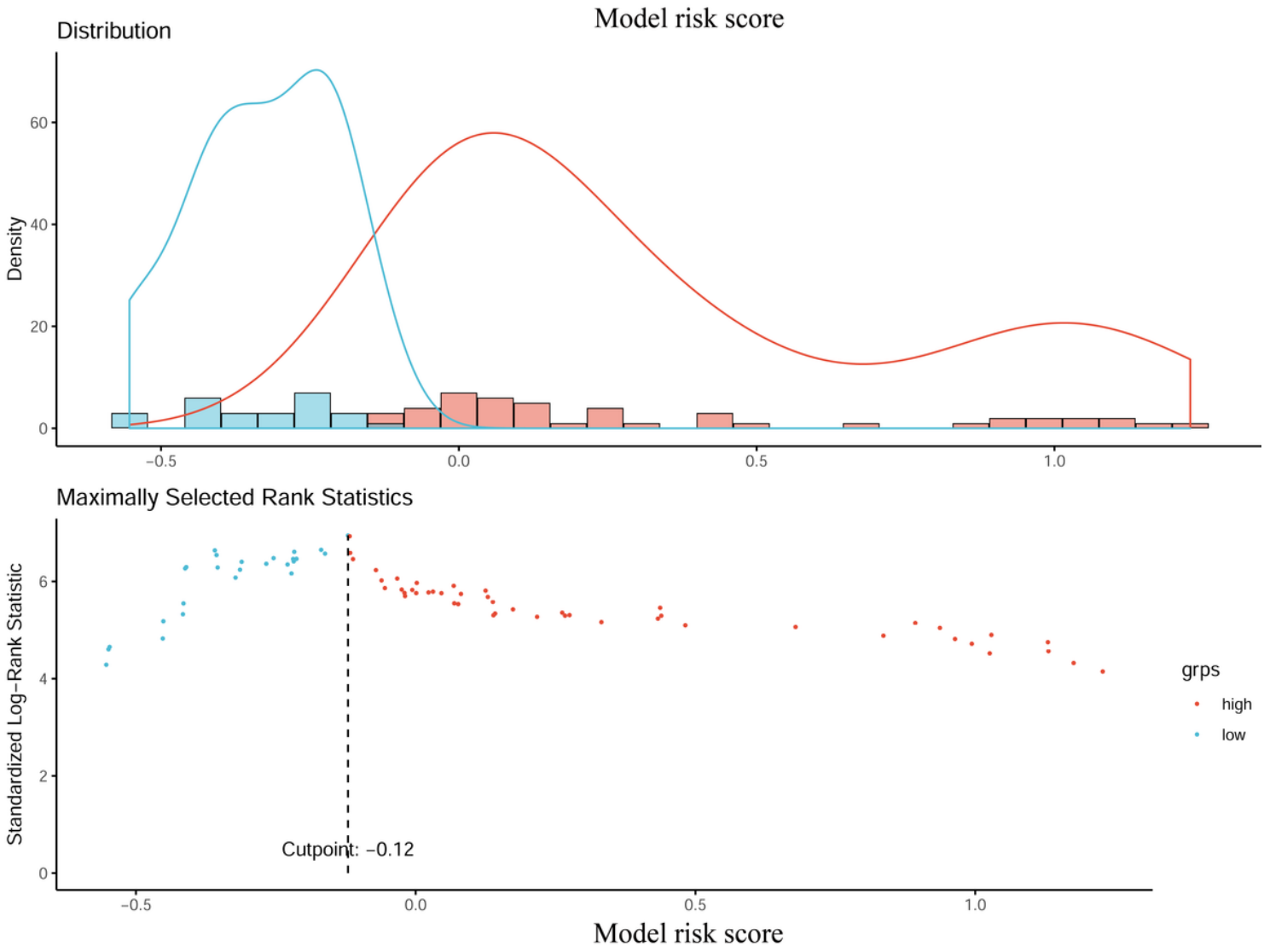

Figure 5

The optimal cut-off value of prognostic model risk score using R package "survival".

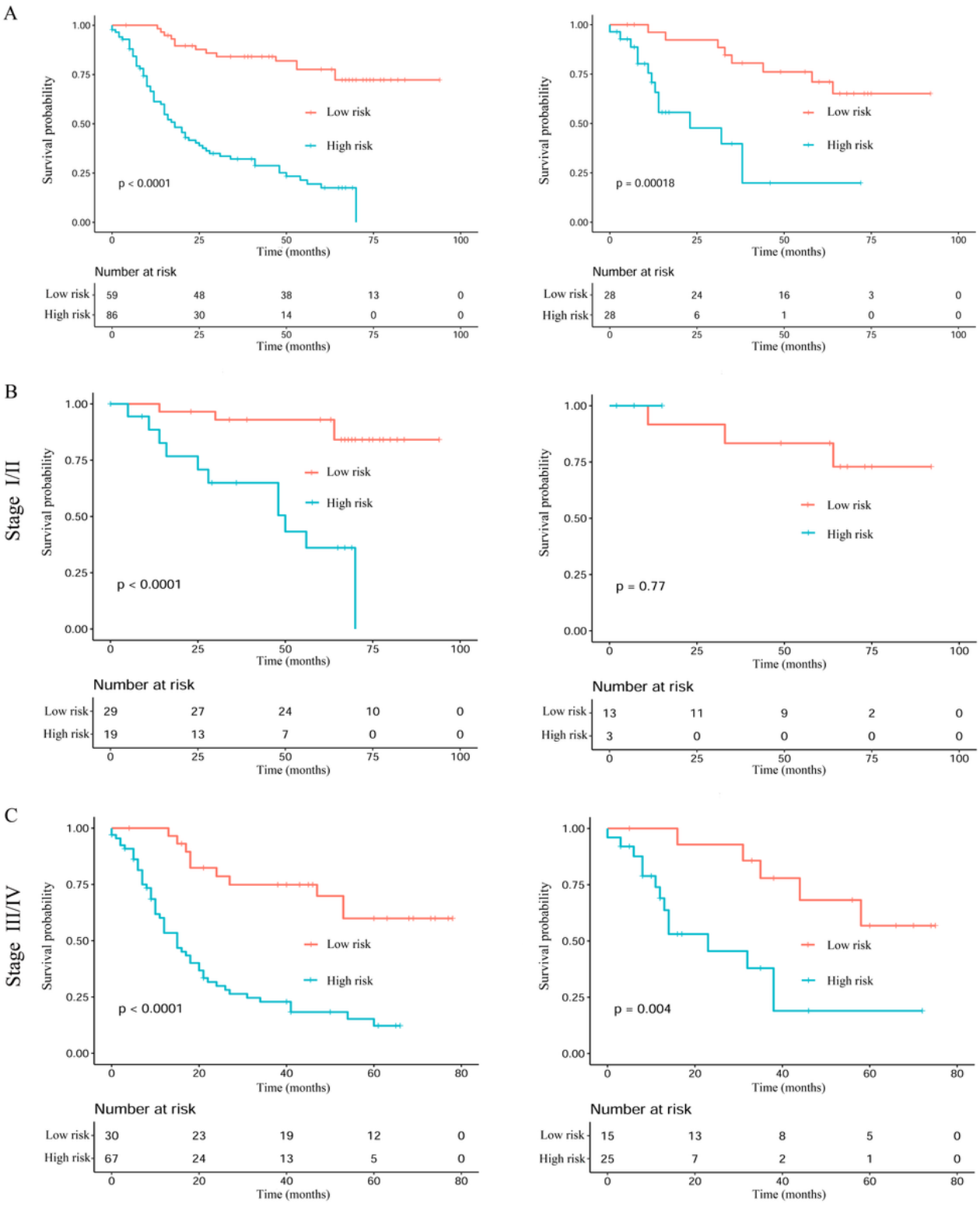

Figure 6

Kaplan–Meier analyses of OS according to the prognostic model risk score classifier in subgroups of NSCLC (HBV+) patients in the training cohort (left) and the validation cohort (right): (A) Total patients; (B) Stage I/II; (C) Stage III/IV.

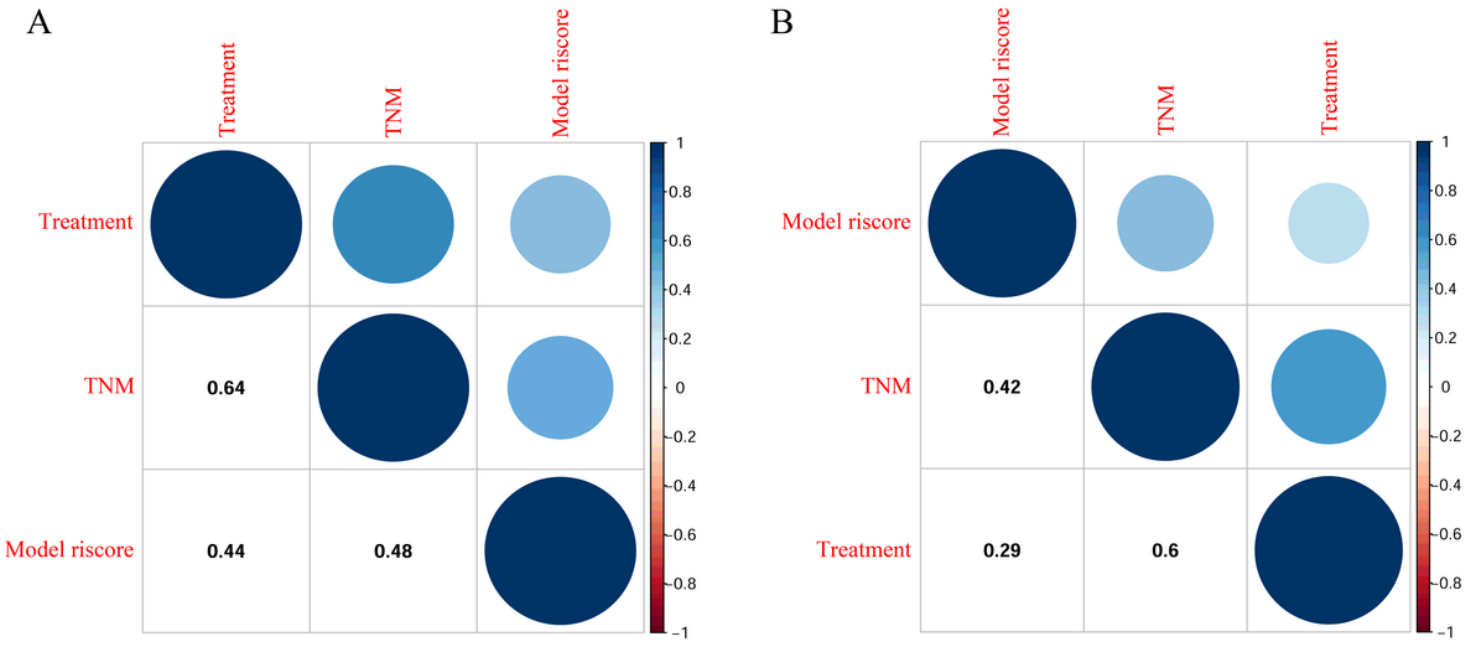

Figure 7

The correlations between the prognostic model and TNM staging or clinical treatment in training cohort (A) and validation cohort (B).