DOI: 10.53469/jcmp.2024.06(12).36

# Prediction of Postoperative Survival Nomogram of Colorectal Cancer Patients based on Lymph Node Classification: a Study based on a SEER Population

**Xuefei Zhang, Lanxin Hu, Junfeng Du, Chen Su, Xiang Xu**

Department of Gastrointestinal Surgery, the First Affiliated Hospital of Chongqing Medical University, Chongqing, China

**Abstract:** ***Background:*** ***Lymph node metastases have a significant impact on the stage, treatment, and prognosis of patients with Colorectal cancer. Lymph node metastasis (LNM), log odds of positive lymph nodes (LODDS), and lymph node ratio (LNR) are independent prognostic factors for colon cancer. We are working to explore a more accurate prediction model and compare it with the staging predictions proposed by the American Joint Committee on Cancer (AJCC).*** ***Methods:*** ***A total of 15,957 patients with colorectal cancer who underwent surgical resection were included in the study, and they were randomly divided into training group (11,169 patients) and validation group (4,788 patients). Univariate Cox analysis, random forest regression analysis and Lasso analysis were performed on the training group. Based on the results of the analysis, the best independent prognostic factors were identified and integrated to construct a nomogram. The concordance index (C-index) and calibration curves were used to evaluate the nomograms of the training group and the validation group. Nomograms were compared to AJCC 8th edition TNM staging system using decision curve analysis (DCA) and area under curve (AUC).*** ***Results:*** ***N, LODDS, and LNR are independent prognostic factors for colorectal cancer. The C-index of nomogram predicting overall survival (OS) is higher than that of the AJCC 8th edition TNM staging system. Decision curve analysis (DCA) and ROC curve suggested that the nomogram was better than AJCC 8th TNM stage in terms of clinical practicability.*** ***Conclusions:*** ***We constructed a nomogram of the prognosis of patients with colorectal cancer, which may help clinicians provide individualized treatment.***

**Keywords:** Nomogram, Log odds of positive lymph nodes (LODDS), Lymph node ratio (LNR), Colorectal cancer (CRC).

## 1. Background

Colorectal cancer (CRC) is the second leading cause of cancer death in the United States, and in China, its incidence is also increasing year by year, and it is currently the third largest among all tumor types, posing a significant threat to people's health [1-2]. The most effective treatment for CRC is based on the TNM staging system developed by the American Joint Committee on Cancer (AJCC) and the International Union Against Cancer (UICC) [3]. In this system, lymph node metastases (LNM) is considered one of the key prognostic indicators to assess overall survival (OS) in colorectal cancer patients [4-7].

In the 8th edition TNM staging system, N is divided into N1 (1~3 LNMs) and N2 (≥4 LNMs). Lymph node ratio (LNR) is the ratio between the number of positive lymph nodes and the total number of lymph nodes detected. LNR has been shown to be an important independent predictor of prognosis in colorectal cancer patients in numerous studies [7-11]. Recently, log odds of positive lymph nodes (LODDS) has begun to attract attention. LODDS, which refers to the logarithm of the ratio of the number of positive lymph nodes to the number of negative lymph nodes, has been included in colorectal cancer studies and is considered a new prognostic factor. Wang et al. [12] found that LODDS staging is a more effective prognostic assessment tool in a prognostic analysis of 24,477 patients with stage III colon cancer. A retrospective study of 115 patients with stage IV colon cancer by Ozawa et al. [13] reached similar conclusions. Other independent prognostic factors, including sex, age, grade, tumor serum biomarkers (carcinoembryonic antigen, CEA), tumor deposits (TDs), and treatment-related factors, have also been shown to play a significant role in personalized survival prediction [14][15].

Nomogram is a tool for building statistical prediction models by integrating key factors in tumor prognosis, which is capable of calculating the probability of clinical events occurring and has been widely used in clinical practice [16]. Compared with traditional TNM staging system or other staging methods, nomograms have shown accuracy in predicting a variety of tumors [17-19].

In this study, a predictive model was developed to assess patient prognosis using the SEER database, taking into account N stage, LNR, LODDS, and other risk factors related to prognosis. This model is designed to provide clinicians with an accurate prognostic assessment tool to improve patients' quality of life and survival expectations.

## 2. Material and Methods

### 2.1 Patients Select

In this study, patients diagnosed with CRC between 2013 and 2015 were screened using version 8.4.3 of the SEER*Stat software. Inclusion criteria included: (1) diagnosis limited to 2013 to 2015; (2) Colorectum is the primary site of the tumor, and the pathological diagnosis is adenocarcinoma; (3) Malignant tumor according to the International Classification of Diseases of Oncology, Third Edition (ICD-O-3); (4) the patient has only one primary tumor; (5) The patient underwent surgical treatment and lymph node dissection; (6) Tumor survival time is at least 1 month. Exclusion criteria include: (1) multiple tumors; (2) non-colorectal primary tumors; (3) Younger than 18 years old or over 80 years old; (4) There is a lack of information on tumor size, grade, and marital status at the time of diagnosis. We updated the TNM staging of patients enrolled in the SEER database according to the AJCC 8th edition criteria. The primary observational endpoint of this study is overall survival (OS), which is the time interval

from the initial diagnosis of CRC to the last follow-up.

The clinical information included in this study included age, sex, tumor site, grade, T stage, N stage, distant metastases (M), total lymph node, LNR, LODDS, tumor size, CEA, TDs, and marital status. X-tile software was used to divide the two continuous variables of LODDS and LNR into three points. LODDS was divided into LODDS0 (≥-2.25, ≤-1.41), LODDS1 (>-1.41, ≤-0.17), LODDS2 (>-0.17, ≤2.08), and LNR0 (≤0.05), LNR1 (> 0.05, ≤ 0.38), LNR2 (>0.38, ≤1).

**2.2 Statistical Analysis**

Patients were randomly assigned to the training and validation groups in a ratio of 7:3. We use X-tile software to divide the continuous variables into categorical variables and determine the optimal cleavage point. Using the chi-square test, we compared the baseline characteristics of patients in the training and validation groups. Next, we used COX univariate regression analysis, random forest regression analysis, and lasso analysis to identify independent prognostic factors. Based on these factors, we construct a multivariate Cox proportional hazards model, and based on this model, a prediction nomogram is established. C-index, area under curve (AUC), calibration curve and decision curve analysis (DCA) were used for internal and external verification. SPSS (version 27.0) and R (version 4.4.1), X-tile (version 3.6.1) were used for statistical analysis. P<0.05 was considered statistically significant.

## 3. Results

A total of 15,957 patients were included in this study. 11,169 (70.0%) patients were randomly assigned to the training group and 4,788 (30.0%) patients were assigned to the validation group. The continuous variables were converted into categorical variables, and the chi-square test was used to compare the differences in clinical information between the training and validation group. There was no significant difference between the two groups (Table1). There were 8570 males (53.71%) and 7387 females (46.29%). The tumors were located in 8378 cases (52.50%) of the right colon, 4929 cases (30.89%) of the left colon, and 2650 cases (16.61%) of the rectum. There were 8227 patients (51.56%%) with N0 stage, 4503 patients (28.22%) with N1 stage, and 3227 patients (20.22%) with N2 stage. 9471 patients (59.35%%) had LNR0 stage, 4855 patients (30.43%%) had LNR1 stage, and 1631 patients (10.22%) had LNR2 stage. There were 7304 patients (45.77%) with LODDS0, 7099 (44.49%) with LODDS1 and 1554 (9.74%) with LODDS2. There were 2626 cases (16.46%) with TDs; There were 6373 cases (39.94%) positive for CEA.

Univariate Cox regression analysis revealed a number of factors significantly associated with overall survival (OS), including age, marital status, tumor site, grade, T stage, N stage, chemotherapy, M, LNR, LODDS, TDs, CEA, Total lymph nodes, and tumor size (P<0.05). However, there was no statistically significant difference between sex in univariate Cox regression analysis(P>0.05).

We further included these statistically significant variables in random forest analysis and evaluated their importance, and the importance ranking of the variables is shown in Figure 2. The analysis found that N, LNR, and LODDS had high importance. Subsequently, we removed the two least important variables from the analysis and included the remaining variables in the LASSO regression analysis. Using the Lasso regression model, we determined the optimal penalty coefficient λ and screened the variables at λ+S to finally identify 12 OS-related factors, including age, tumor size, grade, T stage, N stage, M stage, chemotherapy, CEA, TDs, LNR, LODDS, and tumor size. Based on the results of lasso regression, we constructed a multivariate Cox regression prediction model for OS and presented it in the form of a nomogram. Although the P value of N stage in Cox multivariate regression analysis was greater than 0.05, according to previous studies, N stage had an important impact on the prognosis of tumor survival, so we still included it in the regression model.

The results showed that the C- index of the nomogram was 0.793 (95%CI: 0.786~0.800), and that of the validation group was 0.786 (95% CI: 0.774~0.798), which was better than that of TNM stage (C-index: training group, 0.752, 95%CI, 0.744~0.761; validation group 0.750, 95%CI, 0.737~0.763). The results show that the nomogram has higher predictive power than TNM. The calibration curves of the training group and the validation group were close to 45°, which indicated that the prediction results of the model were in good agreement with the actual results, and the prediction performance of the model was more reliable (Figure 5), which was closer to the real situation. The DCA curves of the training and validation groups (Figure 6) show that our nomogram is superior to TNM staging system. At the same time, the ROC curves of 1-year, 3-year, and 5-year OS in the training group and the validation group were analyzed, and the results showed that the AUC of 1-year, 3-year, and 5-year OS in the training group was 0.836 (95%CI: 0.822~ 0.848), 0.845 (95%CI: 0.836~ 0.852), and 0.832 (95%CI: 0.824~ 0.842), respectively, and the AUC of 1-year, 3-year, and 5-year OS in the validation group were 0.843 (95%) CI: 0.826~ 0.863), 0.840 (95%CI: 0.824~ 0.854), 0.826 (95%CI: 0.815~ 0.839), the larger the AUC, the higher the accuracy of the model prediction (Figure 7).

## 4. Discussion

LNM has a significant impact on the prognosis of CRC patients. Although N stage, LNR and LODDS all showed significant effects on the survival prognosis of CRC patients in univariate regression analysis, random forest analysis and lasso regression analysis, in multivariate regression analysis, N stage was not significantly associated with the survival prognosis of CRC patients. This result is inconsistent with the literature we have previously read. We hypothesize that since the N stage, LNR, and LODDS are all node-based classifications, there are similarities between them, which may lead to the effect of N stage on OS being masked by the LNR and LODDS. At the same time, Ben et al. [20] found a significant correlation between LODDS and LNR. In addition, the prognostic prediction accuracy of N stage was significantly affected by the total number of lymph nodes retrieved. The predictive accuracy of N stage can only be guaranteed when the number of lymph nodes retrieved and examined reaches 12 or more [21].

LODDS and LNR are ratio-based approaches to lymph node assessment, both of which include the total number of lymph nodes dissected and the number of positive lymph nodes, which can overcome the limitations of quantity-based assessment to some extent [10-12]. At present, there is no conclusive jury as to which LNR or LODDS is the best predictor of survival in patients with CRC. The results of Wang et al. [12] suggest that LODDS staging is a better prognostic factor. However, it has also been suggested that LNRs are better suited than LODDS for predicting survival and prognosis in patients with CRC [22]. In our study sample, LNR showed higher predictive value than LODDS, which may be attributed to the study center, the ethnicity composition of the participants, the time frame of the study, and the covariates included in the model. Previous studies have focused on comparing the differences between N stage, LODDS, and LNR, but in our study, these three staging systems are combined into a same prognostic model, which can make better use of the lymph node information obtained during surgery and thus improve the accuracy of patient survival prediction. As our study demonstrates, this nomogram is more effective than TNM stage in terms of prognostic prediction.

TDs have a significant impact on overall survival (OS) as an independent prognostic factor. The presence of TDs usually means that the patient's survival prognosis is poor. In the 8th TNM stag staging system, the N1c classification was introduced to specifically indicate the presence of TDs even in the absence of LNMs, but the effect of TDs on survival prognosis was discarded when lymph node metastases were present. In fact, the presence of TDs should be as prognostic as N stage, and their evaluation should not be limited to the absence of lymph node metastases [24]. Mayo et al. [25] performed different analyses in the same database and showed that the presence of TDs was associated with a lower three-year OS rate in a multivariate model.

In this study, we developed and externally validated a prognostic nomogram predicting the probability of patient survival based on the results of the Cox proportional hazards model. This nomogram has a higher forecast accuracy. Compared with TNM stage, nomograms can integrate multiple prognostic factors to make more personalized predictions for patients.

There are also some limitations to this study. First of all, the SEER database lacks some important clinical information, including complete information about chemotherapy regimens, the presence or absence of immunity and targeted therapy, etc. Secondly, there are differences in the optimal cut-off values for LODDS and LNR in different studies, and if these optimal cut-off values can be determined, their predictive power in clinical practice may be further improved. Then, the SEER data information used in this study was based on colorectal cancer patients in the United States. Therefore, for colorectal cancer patients in China, the applicability of the nomogram needs to be clinically verified.

## 5. Conclusions

In summary, we developed a nomogram model including LODDS and LNR to predict OS in patients with CRC. The validation of the model shows that the model has good discrimination and consistency, which can provide a reliable reference value and theoretical basis for clinicians to carry out personalized diagnosis and treatment models.

**Table 1:** Baseline characteristics of training cohort and external validation cohort

| Variables | Total (n = 15957) | train (n = 11169) | test (n = 4788) | Statistic | *P* |
|---|---|---|---|---|---|
| Age, n(%) | | | | $\chi^2$=1.46 | *0.482* |
| >73 | 2402 (15.05) | 1657 (14.84) | 745 (15.56) | | |
| ≤56 | 5185 (32.49) | 3646 (32.64) | 1539 (32.14) | | |
| 56-73 | 8370 (52.45) | 5866 (52.52) | 2504 (52.30) | | |
| Sex, n(%) | | | | $\chi^2$=1.43 | *0.232* |
| Female | 7387 (46.29) | 5136 (45.98) | 2251 (47.01) | | |
| Male | 8570 (53.71) | 6033 (54.02) | 2537 (52.99) | | |
| Site, n(%) | | | | $\chi^2$=4.39 | *0.111* |
| Left Colon | 4929 (30.89) | 3503 (31.36) | 1426 (29.78) | | |
| Rectum | 2650 (16.61) | 1828 (16.37) | 822 (17.17) | | |
| Right Colon | 8378 (52.50) | 5838 (52.27) | 2540 (53.05) | | |
| Grade, n(%) | | | | $\chi^2$=4.77 | *0.189* |
| Grade I | 1133 (7.10) | 807 (7.23) | 326 (6.81) | | |
| Grade II | 11862 (74.34) | 8253 (73.89) | 3609 (75.38) | | |
| Grade III | 2402 (15.05) | 1701 (15.23) | 701 (14.64) | | |
| Grade IV | 560 (3.51) | 408 (3.65) | 152 (3.17) | | |
| T, n(%) | | | | $\chi^2$=4.25 | *0.236* |
| T1 | 1586 (9.94) | 1080 (9.67) | 506 (10.57) | | |
| T2 | 2511 (15.74) | 1741 (15.59) | 770 (16.08) | | |
| T3 | 8909 (55.83) | 6263 (56.07) | 2646 (55.26) | | |
| T4 | 2951 (18.49) | 2085 (18.67) | 866 (18.09) | | |
| N, n(%) | | | | $\chi^2$=7.98 | *0.018* |
| N0 | 8227 (51.56) | 5690 (50.94) | 2537 (52.99) | | |
| N1 | 4503 (28.22) | 3161 (28.30) | 1342 (28.03) | | |
| N2 | 3227 (20.22) | 2318 (20.75) | 909 (18.98) | | |
| M, n(%) | | | | $\chi^2$=2.62 | *0.106* |
| M0 | 13583 (85.12) | 9474 (84.82) | 4109 (85.82) | | |
| M1 | 2374 (14.88) | 1695 (15.18) | 679 (14.18) | | |
| Chemotherapy, n(%) | | | | $\chi^2$=1.35 | *0.245* |
| No | 8353 (52.35) | 5813 (52.05) | 2540 (53.05) | | |
| Yes | 7604 (47.65) | 5356 (47.95) | 2248 (46.95) | | |
| CEA, n(%) | | | | $\chi^2$=1.64 | *0.201* |
| Negative | 9584 (60.06) | 6672 (59.74) | 2912 (60.82) | | |
| Positive | 6373 (39.94) | 4497 (40.26) | 1876 (39.18) | | |
| TDs, n(%) | | | | $\chi^2$=4.99 | *0.025* |
| No | 13331 (83.54) | 9283 (83.11) | 4048 (84.54) | | |
| Yes | 2626 (16.46) | 1886 (16.89) | 740 (15.46) | | |
| Lymphnode, n(%) | | | | $\chi^2$=0.04 | *0.839* |
| <12 | 1461 (9.16) | 1026 (9.19) | 435 (9.09) | | |
| ≥12 | 14496 (90.84) | 10143 (90.81) | 4353 (90.91) | | |
| LNR, n(%) | | | | $\chi^2$=6.80 | *0.033* |

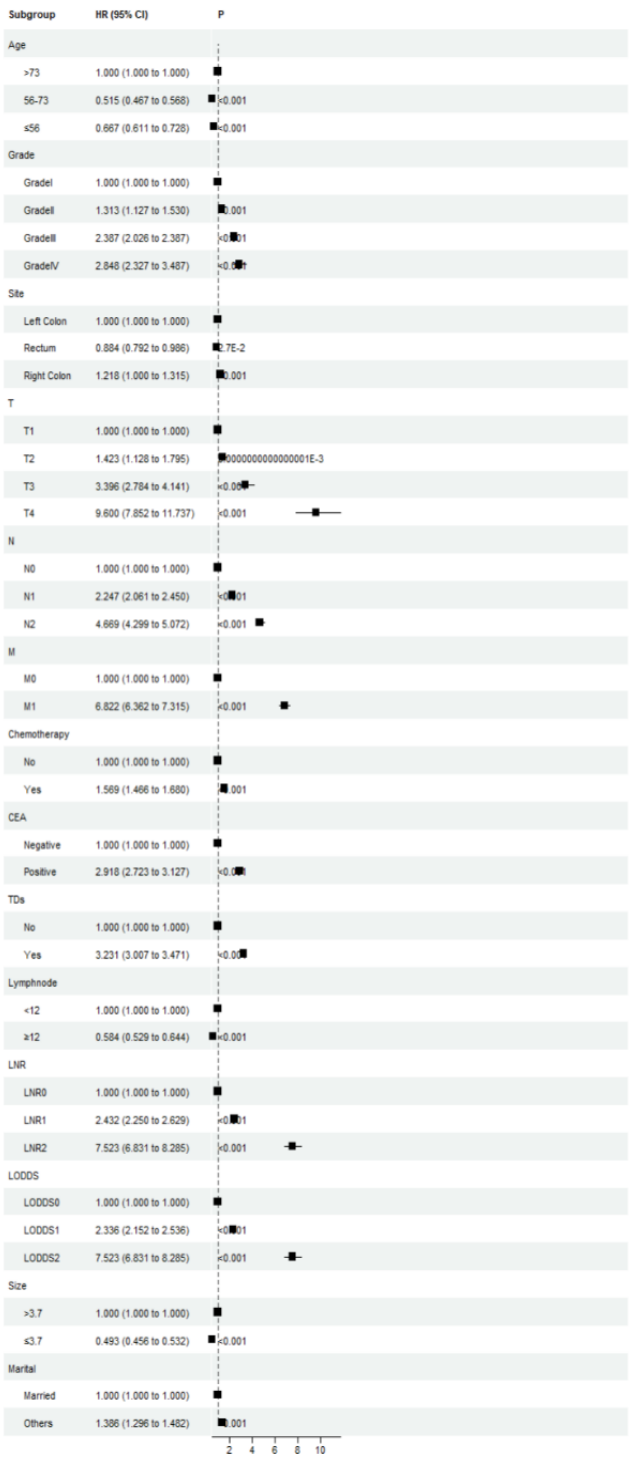
| | | | | |
|---|---|---|---|---|
| LNR0 | 9471 (59.35) | 6567 (58.80) | 2904 (60.65) | |
| LNR1 | 4855 (30.43) | 3422 (30.64) | 1433 (29.93) | |
| LNR2 | 1631 (10.22) | 1180 (10.56) | 451 (9.42) | |
| LODDS, n(%) | | | | $\chi^2$=6.78 *0.034* |
| LODDS0 | 7304 (45.77) | 5056 (45.27) | 2248 (46.95) | |
| LODDS1 | 7099 (44.49) | 4987 (44.65) | 2112 (44.11) | |
| LODDS2 | 1554 (9.74) | 1126 (10.08) | 428 (8.94) | |
| Size, n(%) | | | | $\chi^2$=1.43 *0.232* |
| >3.7 | 9900 (62.04) | 6963 (62.34) | 2937 (61.34) | |
| ≤3.7 | 6057 (37.96) | 4206 (37.66) | 1851 (38.66) | |

Notes: $\chi^2$: Chi-square test; LODDS log odds of positive lymph nodes; LNR lymph node ratio; TDs tumor deposits; CEA carcinoembryonic antigen.

**Table 2:** Univariable analysis and multivariable cox proportional hazards regression analysis

| Variables | Univariable analysis | | Multivariable analysis | |
|---|---|---|---|---|
| | *P* | HR (95%CI) | *P* | HR (95%CI) |
| Age | | | | |
| >73 | | 1.00 (Reference) | | 1.00 (Reference) |
| ≤56 | **<.001** | 0.51 (0.47 ~ 0.57) | **<.001** | 0.46 (0.41 ~ 0.51) |
| 56-73 | **<.001** | 0.67 (0.61 ~ 0.73) | **<.001** | 0.63 (0.58 ~ 0.69) |
| Sex | | | | |
| Female | | 1.00 (Reference) | | |
| Male | **0.007** | 1.10 (1.03 ~ 1.17) | | 1 |
| Site | | | | |
| Left Colon | | 1.00 (Reference) | | 1.00 (Reference) |
| Rectum | **0.027** | 0.88 (0.79 ~ 0.99) | 0.272 | 1.06 (0.95 ~ 1.19) |
| Right Colon | **<.001** | 1.22 (1.13 ~ 1.31) | **<.001** | 1.18 (1.09 ~ 1.27) |
| Grade | | | | |
| GradeI | | 1.00 (Reference) | | 1.00 (Reference) |
| GradeII | **<.001** | 1.31 (1.13 ~ 1.53) | 0.354 | 1.08(0.92 ~ 1.25) |
| GradeIII | **<.001** | 2.39 (2.03 ~ 2.81) | **<.001** | 1.34 (1.13 ~ 1.58) |
| GradeIV | **<.001** | 2.85 (2.33 ~ 3.49) | **<.001** | 1.43 (1.17 ~ 1.76) |
| T | | | | |
| T1 | | 1.00 (Reference) | | 1.00 (Reference) |
| T2 | **0.003** | 1.42 (1.13 ~ 1.79) | 0.166 | 1.18 (0.93~ 1.49) |
| T3 | **<.001** | 3.40 (2.78 ~ 4.14) | **<.001** | 1.85 (1.50 ~ 2.29) |
| T4 | **<.001** | 9.60 (7.85 ~ 11.74) | **<.001** | 3.31 (2.65 ~ 4.14) |
| N | | | | |
| N0 | | 1.00 (Reference) | | 1.00 (Reference) |
| N1 | **<.001** | 2.25 (2.06 ~ 2.45) | **0.104** | 1.15 (0.97 ~ 1.37) |
| N2 | **<.001** | 4.67 (4.30 ~ 5.07) | **0.274** | 1.12(0.91 ~ 1.38) |
| M | | | | |
| M0 | | 1.00 (Reference) | | 1.00 (Reference) |
| M1 | **<.001** | 6.82 (6.36 ~ 7.32) | **<.001** | 3.66 (3.36 ~ 3.98) |
| Chemotherapy | | | | |
| No | | 1.00 (Reference) | | 1.00 (Reference) |
| Yes | **<.001** | 1.57 (1.47 ~ 1.68) | **<.001** | 0.49 (0.45 ~ 0.53) |
| CEA | | | | |
| Negative | | 1.00 (Reference) | | 1.00 (Reference) |
| Positive | **<.001** | 2.92 (2.72 ~ 3.13) | **<.001** | 1.56 (1.44~ 1.68) |
| TDs | | | | |
| No | | 1.00 (Reference) | | 1.00 (Reference) |
| Yes | **<.001** | 3.23 (3.01 ~ 3.47) | **<.001** | 1.35 (1.24 ~ 1.46) |
| LNR | | | | |
| LNR0 | | 1.00 (Reference) | | 1.00 (Reference) |
| LNR1 | **<.001** | 2.43 (2.25 ~ 2.63) | **<.001** | 1.30 (1.12 ~ 1.51) |
| LNR2 | **<.001** | 6.72 (6.15 ~ 7.33) | **<.001** | 2.06 (1.43 ~ 2.97) |
| LODDS | | | | |
| LODDS0 | | 1.00 (Reference) | | 1.00 (Reference) |
| LODDS1 | **<.001** | 2.34 (2.15 ~ 2.54) | **<.001** | 1.34 (1.15 ~ 1.55) |
| LODDS2 | **<.001** | 7.52 (6.83 ~ 8.29) | **0.007** | 1.65 (1.15 ~ 2.36) |
| Lymphnode | | | | |
| <12 | | 1.00 (Reference) | | |
| ≥12 | **<.001** | 0.58 (0.53 ~ 0.64) | | |
| Size | | | | |
| >3.7 | | 1.00 (Reference) | | 1.00 (Reference) |
| ≤3.7 | **<.001** | 0.49 (0.46 ~ 0.53) | **<.001** | 0.84 (0.77 ~ 0.91) |
| Marital | | | | |
| Married | | 1.00 (Reference) | | |
| Others | **<.001** | 1.39 (1.30 ~ 1.48) | | |

Notes: CI, confidence interval; HR, hazard regression LODDS log odds of positive lymph nodes; LNR lymph node ratio; TDs tumor deposits; CEA carcinoembryonic antigen.



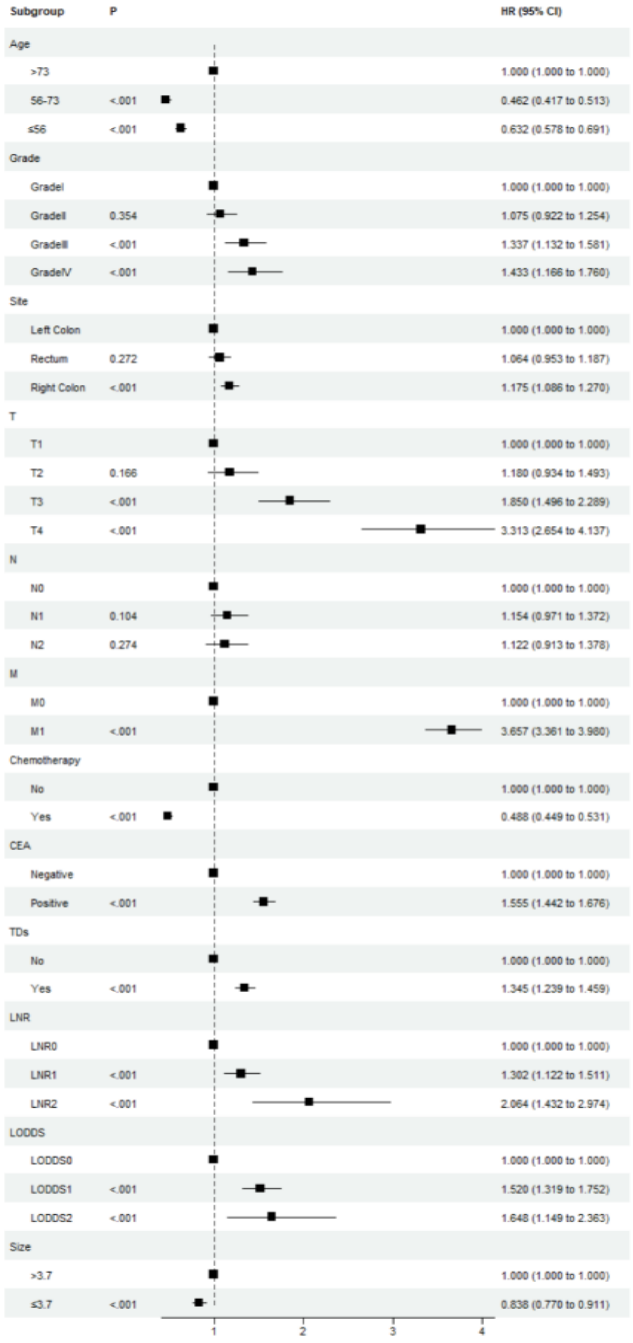



**Figure 1:** Univariable(A) And Multivariable(B) Cox regression analysis and forest plot of prognostic predictors for OS in training cohort. OS, overall survival; LODDS log odds of positive lymph nodes; LNR lymph node ratio;TDs tumor deposits;CEA carcinoembryonic antigen;

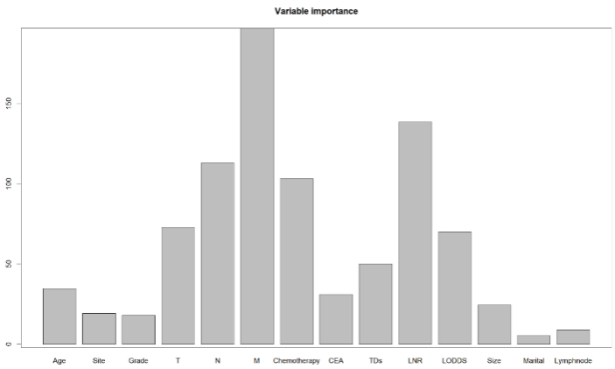



**Figure 2:** The results of variable importance analysis

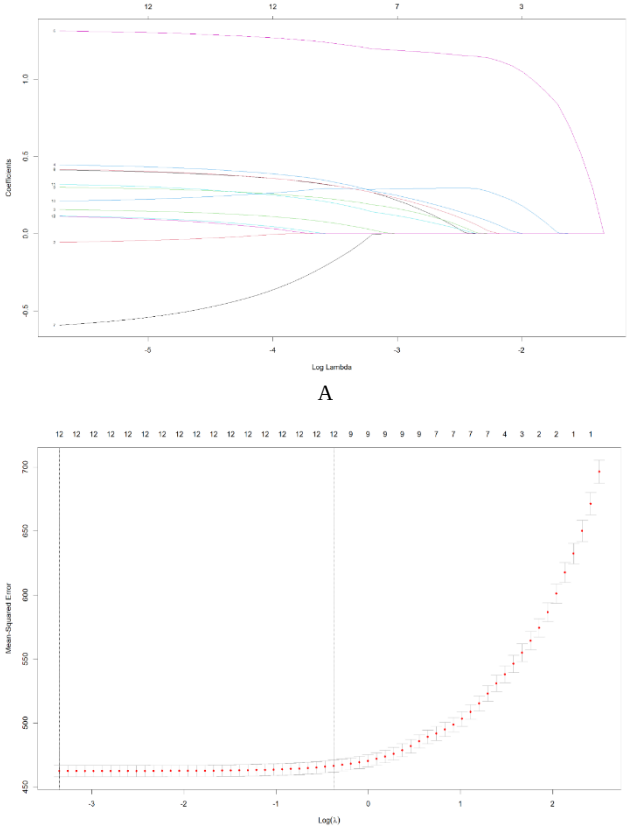



**Figure 3:** Plot of LASSO coefficient profiles of the prognostic predictors for OS in training group (A); Plot of partial likelihood deviance for OS(B));

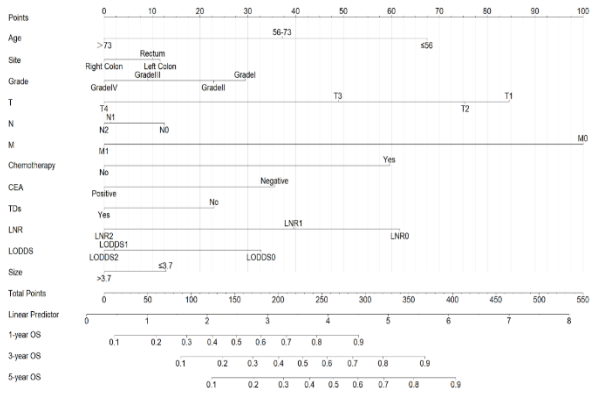
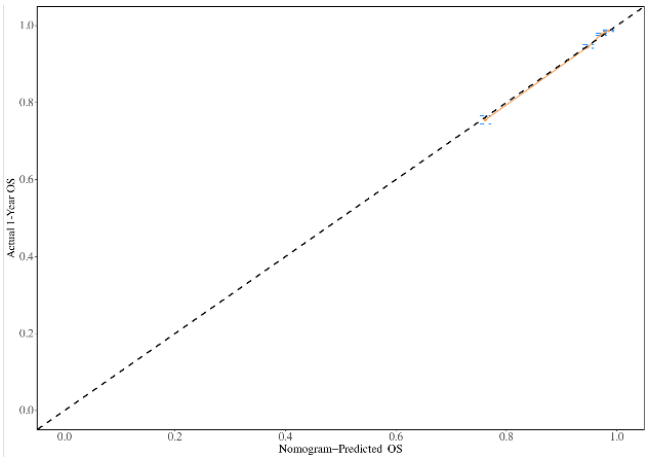
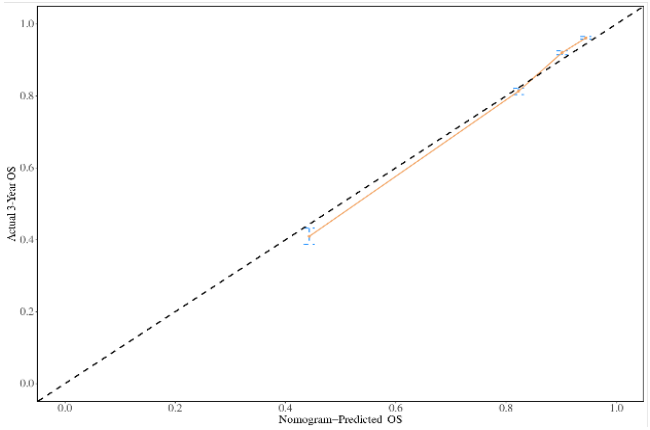
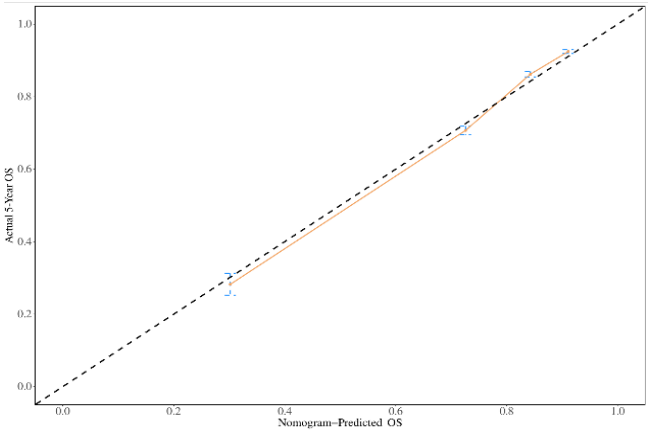



**Figure 4:** Nomograms to predict 1-, 3- and 5-year OS

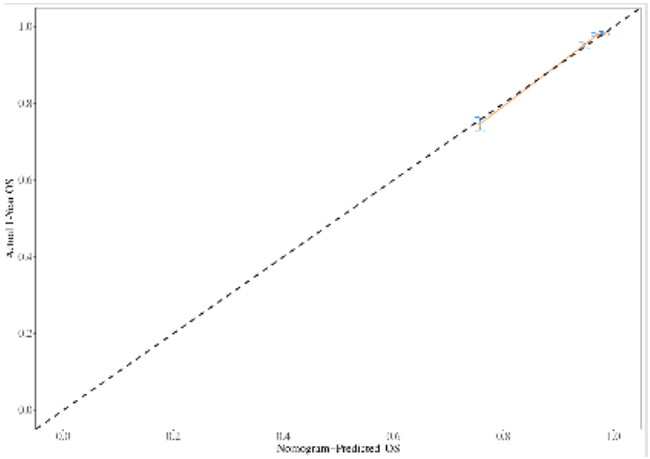




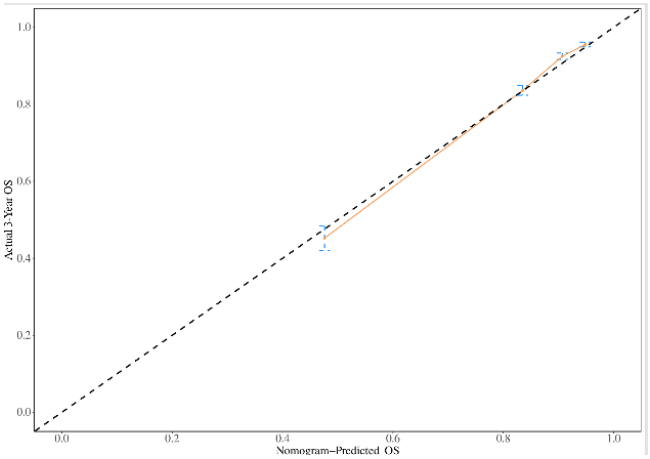

B



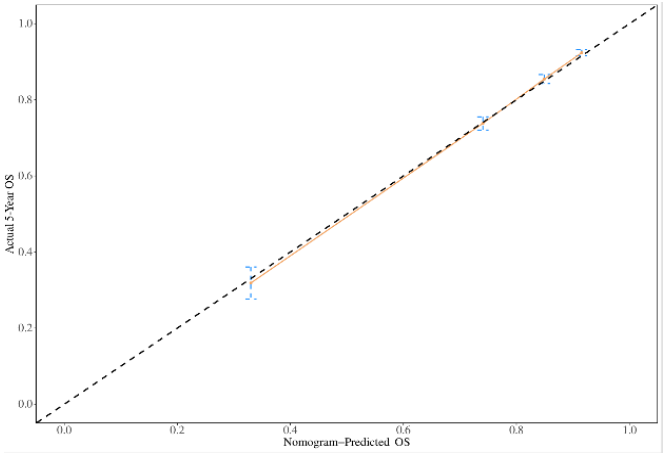

C




D

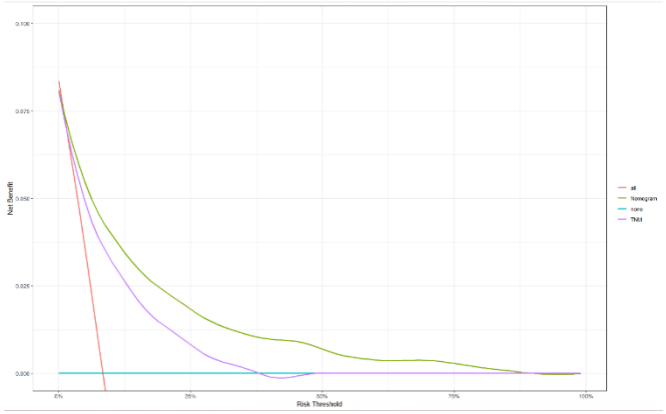



E

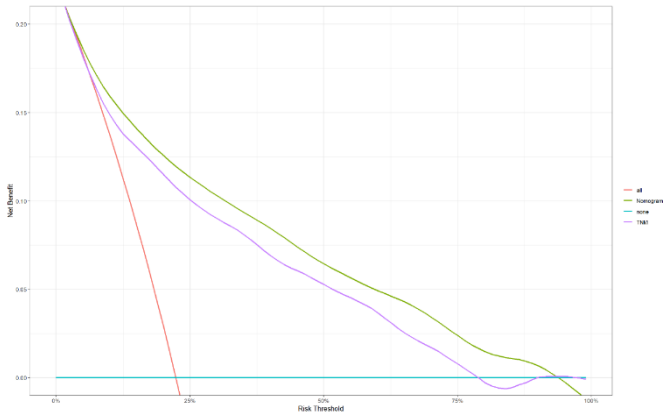



F

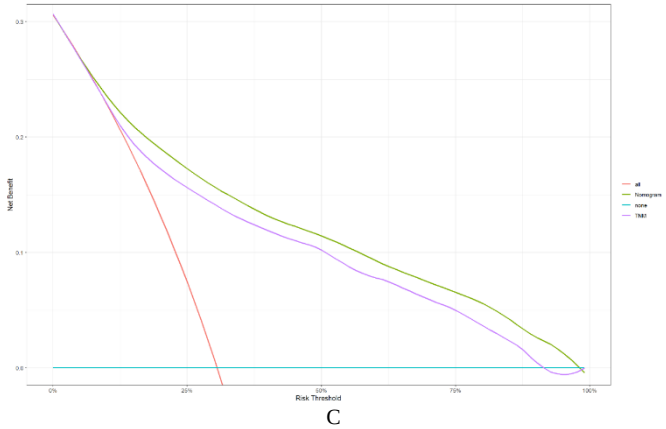
**Figure 5:** Calibration plots of 1-,3-, and 5-year OS in the training group (A-C) and external validation group (D-E); OS, overall survival



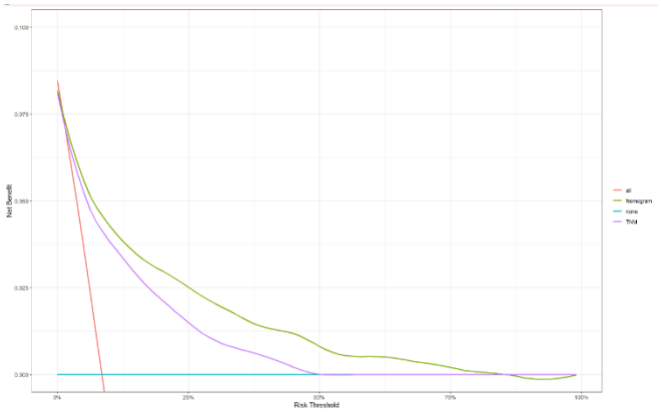

A



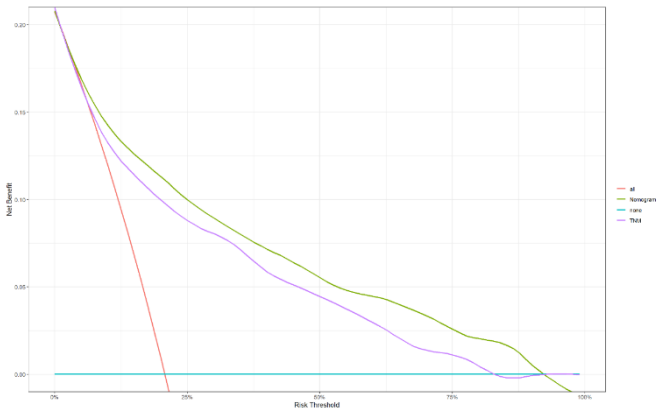

B



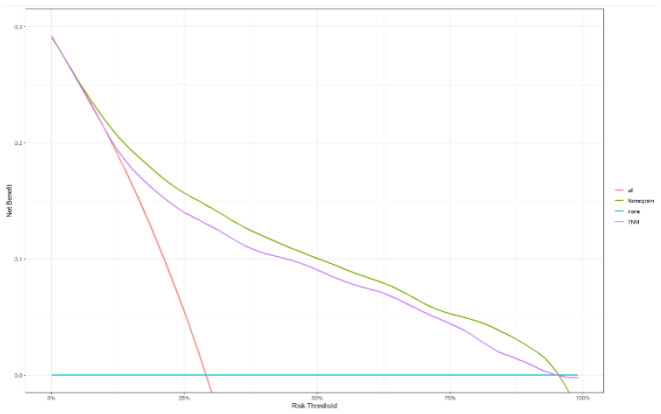

C




D

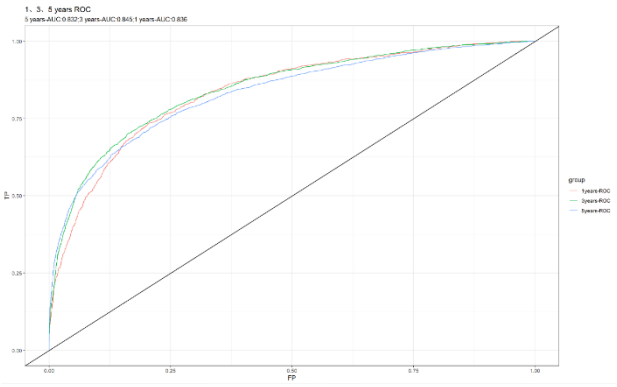



E

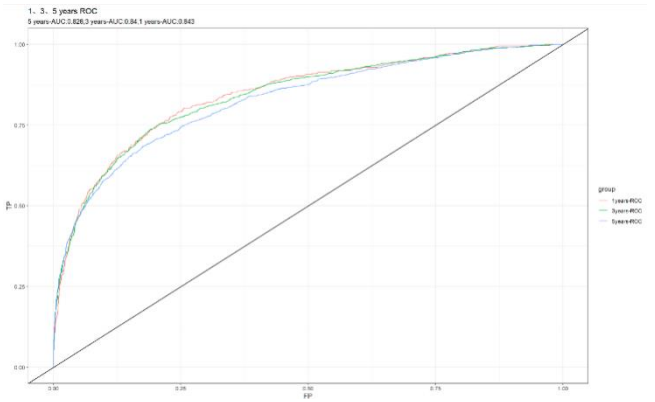



F

**Figure 6:** DCA of TNM and Nomogram for 1-, 3-, and 5-year OS prediction of the training group (A, B, C) and external validation group (D, E, F); DCA, decision curve analysis; TNM, tumor, node, metastasis; OS, overall survival




A




B

**Figure 7:** ROC curves for 1-, 3-, and 5-year OS prediction of the training group (A) and external validation group (B)

## References

[1] ARNOLD M, SIERRA M S, LAVERSANNE M, et al. Global patterns and trends in colorectal cancer incidence and mortality [J]. Gut,2017,66(4):683-691.

[2] Arnold M, Sierra MS, Laversanne M, et al. Global patterns and trends in colorectal cancer incidence and mortality[J]. Gut, 2017, 66(4): 683-691.

[3] Weiser MR. AJCC 8th edition: colorectal cancer. Ann Surg Oncol United States (2018) 25:1454-5.

[4] Kim CW, Kim J, Park Y, Cho DH, Lee JL, Yoon YS, et al. Prognostic implications of extranodal extension in relation to colorectal cancer location. Cancer Res Treat (2019) 51(3):1135-43.

[5] Huh JW, Kim YJ, Kim HR (2010) Ratio of metastatic to resected lymph nodes as a prognostic factor in node-positive colorectal cancer. Ann Surg Oncol 17: 2640-2646.

[6] Edge SB, Byrd DR, Compton CC, Fritz AG, Greene FL, et al. (2010) AJCC cancer staging manual, 7th ed. New York: Springer.

[7] Huh JW, Kim YJ, Kim HR (2010) Ratio of metastatic to resected lymph nodes as a prognostic factor in node-positive colorectal cancer. Ann Surg Oncol 17: 2640-2646.

[8] Qiu HB, Zhang LY, Li YF, Zhou ZW, Keshari RP, et al. (2011) Ratio of metastatic to resected lymph nodes enhances to predict survival in patients with stage III colorectal cancer. Ann Surg Oncol 18: 1568-1574. 7.

[9] Rosenberg R, Engel J, Bruns C, Heitland W, Hermes N, et al. (2010) The prognostic value of lymph node ratio in a population-based collective of colorectal cancer patients. Ann Surg 251: 1070-1078. 8.

[10] Ceelen W, Van Nieuwenhove Y, Pattyn P (2010) Prognostic value of the lymph node ratio in stage III colorectal cancer: a systematic review. Ann Surg Oncol 17: 2847-2855. 9.

[11] Ng M, Roy-Chowdhury S, Lum SS, Morgan JW, Wong JH (2009) The impact of the ratio of positive to total lymph nodes examined and outcome in colorectal cancer. Am Surg 75: 873-876. 10.

[12] Wang J, Hassett JM, Dayton MT, et al. The prognostic supe riority of log odds of positive lymph nodes in stage III colon cancer [J]. J Gastrointest Surg, 2008, 12(10): 1790-1796.

[13] Ozawa T, Ishihara S, Sunami E, et al. Log odds of positive lymph nodes as a prognostic indicator in stage

IV colorectal cancer patients undergoing curative resection [J]. J Surg Oncol, 2015, 111(4):465-471.
[14] Zhang, F., Zhang, Y., Zhao, W., Deng, K., Wang, Z., Yang, C. et al. (2017) Metabolomics for biomarker discovery in the diagnosis, prognosis, survival and recurrence of colorectal cancer: a systematic review. Oncotarget 8, 35460-35472.
[15] Nagtegaal ID, Tot T, Jayne DG, McShane P, Nihlberg A, Marshall HC, et al. Lymph nodes, tumor deposits, and TNM: are we getting better? J Clin Oncol Off J Am Soc Clin Oncol (2011) 29(18):2487-92.
[16] Valentini, V., van Stiphout, R.G.P.M., Lammering, G., Gambacorta, M.A., Barba, M.C., Bebenek, M. et al. (2011) Nomograms for predicting local recurrence, distant metastases, and overall survival for patients with locally advanced rectal cancer on the basis of European randomized clinical trials. J. Clin. Oncol. 29, 3163-3172.
[17] Wen, J., Liu, D., Xu, X., Chen, D., Chen, Y., Sun, L. et al. (2018) Nomograms for predicting survival outcomes in patients with primary tracheal tumors: a large population-based analysis. Cancer Manag. Res. 10, 6843-6856.
[18] Liang, W., Zhang, L., Jiang, G., Wang, Q., Liu, L., Liu, D. et al. (2015) Development and validation of a nomogram for predicting survival in patients with resected non-small-cell lung cancer. J. Clin. Oncol. 33, 861-869.
[19] Wang, Y., Li, J., Xia, Y., Gong, R., Wang, K., Yan, Z. et al. (2013) Prognostic nomogram for intrahepatic cholangiocarcinoma after partial hepatectomy. J. Clin. Oncol. 31, 1188-1195.
[20] Ben Huang, Chen Chen, Mengdong Ni, Shaobo Mo, Guoxiang Cai, Sanjun Cai, Log odds of positive lymph nodes is a superior prognostic indicator in stage III rectal cancer patients: A retrospective analysis of 17,632 patients in the SEER database, International Journal of Surgery, Volume 32, 2016, Pages 24-30, ISSN 1743-9191.
[21] Edge SB, Byrd DR, Compton CC, Fritz AG, Greene FL, et al. (2010) AJCC cancer staging manual, 7th ed. New York: Springer.
[22] Song X, Gao P, Wang ZN, Tong LL, Xu YY, Sun Z, Xing CZ, Xu HM. Which is the most suitable classification for colorectal cancer, log odds, the number or the ratio of positive lymph nodes? PLoS One. 2011;6(12):e28937.
[23] Fan S, Li T, Zhou P, Peng Q, Zhu Y. Development and validation of nomogram combining serum biomarker for predicting survival in patients with resected rectal cancer. Biosci Rep. 2019 Nov 29;39(11): BSR20192636.
[24] Nagtegaal ID, Knijn N, Hugen N, Marshall HC, Sugihara K, Tot T, et al. Tumor deposits in colorectal cancer: improving the value of modern staging-a systematic review and meta-analysis. J Clin Oncol Off J Am Soc Clin Oncol (2017) 35 (10):1119-27.
[25] Mayo E, Llanos AAM, Yi X, Duan S-Z, Zhang L. Prognostic value of tumour deposit and perineural invasion status in colorectal cancer patients: a SEER-based population study. Histopathology (2016) 69(2): 230-8.