Retrospective Study

Development and validation of a prognostic nomogram for colorectal cancer after surgery

Bo-Wen Li, Xiao-Yu Ma, Shuang Lai, Xin Sun, Ming-Jun Sun, Bing Chang

ORCID number: Bo-Wen Li 0000-0001-9556-5546; Xiao-Yu Ma 0000-0001-7131-6940; Shuang Lai 0000-0001-8196-8192; Xin Sun 0000-0003-4381-0255; Ming-Jun Sun 0000-0003-3676-3651; Bing Chang 0000-0003-1965-5827.

Author contributions: Li BW performed the research and wrote the paper; Ma XY supervised the report; Lai S and Sun X contributed to the analysis; Sun MJ and Chang B proposed the idea and clinical advice.

Supported by Science and Technology Support Program of Shenyang, No. 20-205-4-094.

Institutional review board statement: The experimental data are from the Surveillance, Epidemiology, and End Results (SEER) database, not clinical cases from any medical institutions. Therefore, our research does not need to be approved by an ethics committee.

Informed consent statement: The experimental data are from the Surveillance, Epidemiology, and End Results (SEER) database, not clinical cases from any medical institutions. Because the data is anonymous and the patients’ personal privacy information is not available, informed consent is not required.

Abstract

BACKGROUND
A nomogram is a diagram that aggregates various predictive factors through multivariate regression analysis, which can be used to predict patient outcomes intuitively. Lymph node (LN) metastasis and tumor deposit (TD) conditions are two critical factors that affect the prognosis of patients with colorectal cancer (CRC) after surgery. At present, few effective tools have been established to predict the overall survival (OS) of CRC patients after surgery.

AIM
To screen out suitable risk factors and to develop a nomogram that predicts the postoperative OS of CRC patients.

METHODS
Data from a total of 3139 patients diagnosed with CRC who underwent surgical removal of tumors and LN resection from 2010 to 2015 were collected from the Surveillance, Epidemiology, and End Results program. The data were divided into a training set (n = 2092) and a validation set (n = 1047) at random. The Harrell concordance index (C-index), Akaike information criterion (AIC), and area under the curve (AUC) were used to assess the predictive performance of the N stage from the American Joint Committee Cancer tumor-node-metastasis classification, LN ratio (LNR), and log odds of positive lymph nodes (LODDS). Univariate and multivariate analyses were utilized to screen out the risk factors significantly correlating with OS. The construction of the nomogram was based on Cox regression analysis. The C-index, receiver operating characteristic (ROC) curve, and calibration curve were employed to evaluate the discrimination and prediction abilities of the model. The likelihood ratio test was used to compare the sensitivity and specificity of the final model to the model with the N stage alone to
evaluate LN metastasis.

RESULTS
The predictive efficacy of the LODDS was better than that of the LNR based on the C-index, AIC values, and AUC values of the ROC curve. Seven independent predictive factors, namely, race, age at diagnosis, T stage, M stage, LODDS, TD condition, and serum carcinoembryonic antigen level, were included in the nomogram. The C-index of the nomogram for OS prediction was 0.8002 (95%CI: 0.7839-0.8165) in the training set and 0.7864 (95%CI: 0.7604-0.8124) in the validation set. The AUC values of the ROC curve predicting the 1-, 3-, and 5-year OS were 0.846, 0.841, and 0.825, respectively, in the training set and 0.823, 0.817, and 0.835, respectively, in the validation test. Great consistency between the predicted and actual observed OS for the 1-, 3-, and 5-year OS in the training set and validation set was shown in the calibration curves. The final nomogram showed a better sensitivity and specificity than the nomogram with N stage alone for evaluating LN metastasis in both the training set (-4668.0 vs -4688.3, P < 0.001) and the validation set (-1919.5 vs -1919.8, P < 0.001) through the likelihood ratio test.

CONCLUSION
The nomogram incorporating LODDS, TD, and other risk factors showed great predictive accuracy and better sensitivity and specificity and represents a potential tool for therapeutic decision-making.

Core Tip: The Surveillance, Epidemiology, and End Results (SEER) program has provided material and data support for evidence-based clinical studies. At present, few studies have concentrated on developing a predictive model for the outcomes of colorectal cancer (CRC) after surgery. We developed a nomogram to predict the probability of overall survival at different times in patients with CRC based on the SEER database. Compared with the N staging from the American Joint Committee on Cancer tumor-node-metastasis classification, the nomogram incorporating the log odds of positive lymph nodes and tumor deposit in this study showed better sensitivity and specificity.

INTRODUCTION
Colorectal cancer (CRC) is a severe threat to human health, and the mortality of CRC ranks second among the causes of cancer death in the United States[1]. In 2018, the global newly diagnosed CRC cases accounted for 10.2% of the total newly diagnosed cancer cases, and mortality from CRC accounted for 9.2% of the total cancer deaths[2].

The main treatment for CRC is tumor resection. Lymph node (LN) metastasis is one of the most important factors affecting the prognosis of CRC after surgery. Thus, many systems have been proposed to evaluate the LN metastasis conditions. At present, the most widely accepted LN staging system is the N staging system from the American Joint Committee on Cancer (AJCC) tumor-node-metastasis (TNM) classification, which is based on the absolute number of positive lymph nodes (PLNs). However, the N1c stage is an exception; it is defined as a tumor deposit (TD) without PLNs. According to the 8th edition of the AJCC TNM classification, N0 is defined as no
LN metastasis. N1 is defined as 1-3 regional LN metastases and TD without LN positivity. N2 is defined as more than four regional LN metastases, but the stage of the patients who have LN metastases with and without TD is not distinguished clearly[3]. This N staging system requires at least 12 LNs to be excised and examined histopathologically to obtain a reliable result[4], and it is influenced by the number of resected lymph nodes (RLNs), which can cause staging migration.

The LN ratio (LNR) is defined as the ratio of the number of PLNs to RLNs. Some researchers believe that the LNR has a stronger prognostic impact on patients with colon cancer[5], and that it is accurate for predicting the survival of patients with stage II-III rectal adenocarcinoma with limited RLNs[6]. Additionally, the LNR has shown great predictive value in other cancers, such as lung cancer[7] and gastric cancer[8]. However, when the LNR is near 0 and 1, it cannot accurately evaluate the prognosis of patients with CRC[6].

Log odds of positive lymph nodes (LODDS) is defined as the logarithm of the ratio between the number of PLNs and the negative LNs after the LNs are examined. Persian et al[9] suggested that LODDS had a strong predictive ability for patient prognosis, and it was less influenced by the number of LNs resected and examined[9].

Another adverse prognostic factor about LN condition for the overall survival (OS) of patients with CRC is TD. TD is defined as a discrete nodule of cancer in the pericolic/perirectal fat or in the adjacent mesentery (mesocolic or rectal fat) within the lymph drainage area of the primary carcinoma, without any identifiable LN tissue or any identifiable vascular structure[5]. Therefore, TD is necessary to act as another factor to evaluate the LN condition of patients with CRC.

In this study, we used Harrell concordance index (C-index), Akaike information criterion (AIC), and area under the curve (AUC) to identify the most suitable system to evaluate LN metastasis for CRC patients after surgery. Then, we used univariate and multivariate analyses to choose the potential risk factors for CRC, and we developed a nomogram with these risk factors. We assessed the predictive efficacy of the final model and tested it against the validation set. Finally, we compared the nomograms with the N stage alone and with the TD and LODDS instead through the likelihood ratio test.

**MATERIALS AND METHODS**

**Data source and eligibility criteria**

We collected CRC cases from the Surveillance, Epidemiology, and End Results (SEER) database of the American Cancer Institute (http://seer.cancer.gov/). The inclusion criteria were all patients who were histopathologically diagnosed with CRC from 2010 to 2015 and underwent tumor resection surgery. The exclusion criteria included: (1) Unknown age at diagnosis; (2) Unknown race (according to the SEER database, we divided the race into three classifications, white, black, and other; and the American Indian/Alaska Native and Asian/Pacific Islander were included in the other group); (3) Uncertain pathological grade or stage; (4) Uncertain TNM stage information; (5) Unknown tumor size; (6) Unknown extent of LN resection; (7) Undetermined carcinembryonic antigen (CEA) condition; and (8) Uncertain TD condition. Finally, a total of 3139 patients were included in the study. The patients were randomly divided into a training cohort and a validation cohort using R software.

**Statistical analysis**

Categorical variables and continuous variables were compared by chi-square and t-tests, respectively. The optimal cutoff values of the LODDS and LNR were calculated with maxstat function in R in the training set. Univariate and multivariate statistical analyses were used for filtering out the possible hazards of CRC. The risk factors with \( P < 0.05 \) after univariate analysis were included into multivariate analysis. The C-index and AIC were calculated to assess the predictive efficacy of the LODDS, LNR, and N stage for LN metastasis condition. The receiver operating characteristic (ROC) curve and AUC were used to assess the accuracy of the N stage, LODDS, and LNR for the OS of CRC patients 1, 3, and 5 years after surgery. The accuracy of the nomogram was analyzed by the C-index and AUC. A calibration curve was built to illustrate the consistency between the prediction by the nomogram and the actual condition for OS at 1, 3, and 5 years after surgery. Likelihood ratio tests were used to compare the final nomogram and the nomogram with N stage only to evaluate LN metastasis. All statistical analyses were performed with R software in version 4.0.3 using 0.05 as the boundary of statistical significance.
**LNR system**
LNR was defined as the ratio between the PLNs and the total number of retrieved LNs. We grouped the LNR by its optimal cutoff value.

**LODDS system**
The formula of LODDS was $\log \left( \frac{\text{the number of PLNs} + 0.05}{\text{the number of negative nodes} + 0.05} \right)$. In our study, LODDS was grouped into two groups by its optimal cutoff value.

**RESULTS**

**Patient characteristics**
For this research, a total of 3139 patients were included and randomly divided into the training set ($n = 2092, 66.7\%$) and validation set ($n = 1047, 33.3\%$). The detailed clinical data are listed in Table 1. There was no significant difference in age at diagnosis, race, N stage, M stage, serum CEA level, TD, grade, tumor size, LNR, LODDS, PLN, or RLN between the training set and validation set ($P < 0.05$).

**Optimal cutoff value of LODDS and LNR**
We divided the LODDS and LNR into two groups based on their optimal cutoff values calculated with the maxstat function in R. The optimal cutoff values of LODDS and LNR were -0.6504 and 0.1795, respectively.

**Independent prognostic factors for OS**
Univariate analysis indicated that sex, race, T stage, N stage, grade, LODDS, LNR, M stage, tumor size, age, serum CEA level, and TD were all related to OS ($P < 0.05$). The C-indexes of LODDS, LNR, and N stage were 0.6497, 0.6494, and 0.6712, respectively (Table 3). The AICs of N stage, LODDS, and LNR were 9906.03, 9887.95, and 9890.20, respectively. The AUCs predicted that the 1-year, 3-year, and 5-year OS for N stage were 0.695, 0.702, and 0.696, respectively; those for LODDS were 0.690, 0.673, and 0.660, respectively; and those for LNR were 0.689, 0.673, and 0.659, respectively (Table 3). In terms of multivariate Cox analysis, we found that age at diagnosis, race, M stage, T stage, serum CEA level, TD, and LODDS were all independent risk factors ($P < 0.05$, Table 2).

**Prognostic nomogram for OS**
All independent risk factors after screening were integrated into the final predicted nomogram (Figure 1). The predicted C-index of the nomogram for OS prediction was 0.8002 (95%CI: 0.7839-0.8165) in the training set and 0.7864 (95%CI: 0.7604-0.8124) in the validation set. The AUC values of the ROC curve predicting the 1-, 3-, and 5-year OS were 0.846, 0.817, and 0.835 in the validation test (Figure 2A-F). The consistency between the predicted condition by the nomogram and the observed condition for OS at 1, 3, and 5 years after surgery is shown in the calibration curves in the training set (Figure 3A-C) and the validation set (Figure 3D-F).

**Comparison of the nomogram with the N stage**
To compare the predictive efficacy between the final nomogram and the nomogram with the N stage alone, we eliminated the LODDS system and TD condition and incorporated the N stage into the nomogram instead to evaluate the LN metastasis. The N stage was identified to be an independent risk factor through multivariate analysis. The log likelihood values of the final model and the model with N stage were -4668.0 and -4688.3 in the training set ($P < 0.001$) and -1919.5 and -1919.8 in the validation set ($P < 0.001$), respectively.

**DISCUSSION**
A nomogram can intuitively display the risk by synthesizing and explaining the relative importance of various predictive factors. It has been used to predict the outcomes of many diseases[10,11].
| Demographic or characteristic | All subjects (n = 3139) | Training set (n = 2902) | Validation set (n = 1047) | P value |
|-------------------------------|-------------------------|--------------------------|---------------------------|---------|
| Sex                           |                         |                          |                           | 0.0239* |
| Male                          | 1560                    | 1070                     | 490                       |         |
| Female                        | 1579                    | 1022                     | 557                       |         |
| Age at diagnosis, yr          |                         |                          |                           | 0.7318  |
| 20-55                         | 853                     | 570                      | 283                       |         |
| 55-65                         | 722                     | 469                      | 253                       |         |
| 65-75                         | 736                     | 497                      | 239                       |         |
| > 75                          | 828                     | 556                      | 272                       |         |
| Race                          |                         |                          |                           | 0.7521  |
| Black                         | 378                     | 254                      | 124                       |         |
| White                         | 2445                    | 1622                     | 823                       |         |
| Other (American Indian/Alaska Native, Asian/Pacific Islander) | 316 | 216 | 100 |         |
| T stage                       |                         |                          |                           | 0.0141* |
| Tis                           | 36                      | 27                       | 9                         |         |
| T1                            | 458                     | 306                      | 152                       |         |
| T2                            | 499                     | 329                      | 170                       |         |
| T3                            | 1522                    | 1101                     | 421                       |         |
| T4                            | 624                     | 419                      | 205                       |         |
| N stage                       |                         |                          |                           | 0.3186  |
| N0                            | 1633                    | 1084                     | 549                       |         |
| N1                            | 876                     | 573                      | 303                       |         |
| N2                            | 630                     | 435                      | 195                       |         |
| M stage                       |                         |                          |                           | 0.7900  |
| M0                            | 2713                    | 1811                     | 902                       |         |
| M1                            | 426                     | 281                      | 145                       |         |
| Serum CEA level               |                         |                          |                           | 0.7262  |
| Normal                        | 1838                    | 1230                     | 608                       |         |
| Elevated                      | 1301                    | 862                      | 439                       |         |
| Tumor deposit                 |                         |                          |                           | 0.9484  |
| Negative                      | 2698                    | 1797                     | 901                       |         |
| Positive                      | 441                     | 295                      | 146                       |         |
| Grade                         |                         |                          |                           | 0.6182  |
| I                             | 279                     | 185                      | 94                        |         |
| II                            | 2096                    | 1405                     | 691                       |         |
| III                           | 573                     | 383                      | 190                       |         |
| IV                            | 191                     | 119                      | 72                        |         |
| Tumor size                    |                         |                          |                           | 0.0809  |
| ≤ 30 mm                       | 876                     | 597                      | 279                       |         |
| 30-50 mm                      | 934                     | 642                      | 292                       |         |
This research indicated that age at diagnosis, race, T stage, LODDS, M stage, serum CEA level, and TD were independent prognostic factors of CRC after radical tumor resection. Based on these prognostic factors, a nomogram was drawn to predict OS after surgery in CRC patients. The nomogram was developed based on these prognostic factors and was used to predict the 1-year, 3-year, and 5-year OS of CRC patients after surgery. This nomogram showed better sensitivity and specificity than the nomogram with T stage alone in both the training set and the validation set.

The evaluation of LN metastasis for CRC patients has important staging, prognostic, and therapeutic implications. At present, the most widely accepted system for evaluating LN metastasis is N staging from the AJCC TNM classification. This system is mainly based on the absolute number of LNs and is influenced by the number of LNs resected. It has been verified that colon cancer survival is positively associated with the number of LNs analyzed[12]. Some scholars have indicated that the predictive ability of PLNs is based on a sufficient number of RLNs[13]. Thus, the determination of a suitable number of RLNs is important to determine N staging and to predict patient outcomes. However, the recommendations for the ideal number of RLNs are different in different studies. Tsai et al.[14] recommended that the suitable number of RLNs is 18. They showed that the 5-year OS rate of CRC patients who had 18 or more LNs examined was higher than those who had fewer than 18 lymph nodes examined[14]. Swanso et al.[15] believed that a minimum of 13 RLNs were necessary to accurately predict the survival of T3N0 CRC patients[15]. In clinical practice, it is difficult to ensure that all patients have enough RLNs due to the individual differences among different patients as well as differences in surgical methods and the skills of the surgeons.

The factors mentioned above undoubtedly influence the determination and predictive accuracy of N stage, which is named ‘stage migration’. In addition, according to the AJCC TNM classification, the TD status of patients with regional LN metastasis is still unclear.

To reduce the stage migration caused by the N stage, some experts suggest that both the PLN and RLN should be taken into account in the LN evaluation system[13]. The LNR and LODDS were proposed as prognostic factors for CRC[16-19]. The formula to calculate LNR is PLN/RLN. When the number of RLNs is inadequate, the LNR compensates for the under-staging of the N stage and provides a better estimation of prognosis than the N stage[20]. One study found that the 5-year recurrence-free survival rates of CRC patients in different LNR groups but in the same N stage were significantly different, but the differences in the 5-year recurrence-free survival rates of patients in different N stages in the same LNR groups were not significantly different[21], which indicated a better evaluation ability of the LNR than the N stage and reduced stage migration. However, LNR has its own limitations. The LNR cannot eliminate stage migration completely. When the LNR value is near 0, it is equal to the controversial N stage and cannot predict the outcome accurately[6,22].

LODDS is another novel system to evaluate LN metastasis. The formula to calculate LODDS is log[(PLN + 0.05)/(RLN−PLN + 0.05)]. This formula avoids many patients with 0 PLN from having a LODDS of 0 and thus further eliminates stage migration[22]. LODDS can discriminate patients with different numbers of RLNs in the N0 stage (LNR = 0). The OS rates of LN-negative (LNR = 0) patients in different LODDS groups were found to be significantly different, which indicated that the LODDS system was more accurate than the LNR system[19]. Previous studies also verified the better predictive ability of LODDS relative to N stage and LNR in gallbladder cancer and cervical cancer[23,24]. A single-center analysis of 323 CRC patients showed that LODDS had a better predictive value than the N stage and LNR systems[25].
| Variable | Univariate analysis | Multivariate analysis |
|----------|---------------------|-----------------------|
|          | HR (95%CI)          | P value               | HR (95%CI)          | P value               |
| Sex      |                     |                       |                      |                       |
| Female   | Reference            | Reference             |                       |                       |
| Male     | 0.82 (0.70-0.95)     | 0.008<sup>a</sup>     | 1.01 (0.85-1.16)     | 0.9277                |
| Age at diagnosis, yr |     |                       |                      |                       |
| 20-55    | Reference            | Reference             |                       |                       |
| 55-65    | 1.12 (0.88-1.43)     | 0.339                 | 1.38 (1.1380-1.6278) | 0.0095<sup>b</sup>    |
| 65-75    | 1.17 (0.93-1.49)     | 0.185                 | 1.47                 | 0.0017<sup>b</sup>    |
| > 75     | 2.69 (2.20-3.30)     | < 0.001<sup>c</sup>   | 3.95 (3.7336-4.1610) | < 0.001<sup>c</sup>   |
| Race     |                     |                       |                      |                       |
| Black    | Reference            | Reference             |                       |                       |
| White    | 0.97 (0.78-1.21)     | 0.799                 | 0.72 (0.4852-0.9513) | 0.0054<sup>d</sup>    |
| Other (American Indian/Alaska Native, Asian/Pacific Islander) | 0.68 (0.48-0.95) | 0.023<sup>a</sup> | 0.64 (0.2921-0.9791) | 0.0097<sup>c</sup>    |
| T stage  |                     |                       |                      |                       |
| Tis      | 1.40 (0.50-3.95)     | 0.524                 | 1.50 (0.4674-2.5395) | 0.4405                |
| T1       | Reference            | Reference             |                       |                       |
| T2       | 1.75 (1.16-2.63)     | 0.00<sup>b</sup>      | 1.21 (0.7887-1.6327) | 0.3745                |
| T3       | 3.24 (2.30-4.57)     | < 0.001<sup>c</sup>   | 1.56 (1.1640-1.9478) | 0.0271<sup>c</sup>    |
| T4       | 8.30 (5.85-11.76)    | < 0.001<sup>c</sup>   | 2.58 (2.1710-3.0036) | < 0.001<sup>c</sup>   |
| N stage  |                     |                       |                      |                       |
| N0       | Reference            | Reference             |                       |                       |
| N1       | 1.85 (1.53-2.24)     | < 0.001<sup>c</sup>   | 1.16 (0.9433-1.3720) | 0.1806                |
| N2       | 4.49 (3.76-5.36)     | < 0.001<sup>c</sup>   | 1.10 (0.7504-1.4569) | 0.5841                |
| M stage  |                     |                       |                      |                       |
| M0       | Reference            | Reference             |                       |                       |
| M1       | 5.95 (5.06-6.99)     | < 0.001<sup>c</sup>   | 3.27 (3.0738-3.4599) | < 0.001<sup>c</sup>   |
| Serum CEA level |     |                       |                      |                       |
| Elevated | Reference            | Reference             |                       |                       |
| Normal   | 0.43 (0.37-0.50)     | < 0.001<sup>c</sup>   | 0.78 (0.6157-0.9438) | 0.0030<sup>d</sup>    |
| Tumor deposit |     |                       |                      |                       |
| Negative | Reference            | Reference             |                       |                       |
| Positive | 3.57 (3.02-4.23)     | < 0.001<sup>c</sup>   | 1.44 (1.2366-1.6385) | 0.0004<sup>c</sup>    |
| Grade    |                     |                       |                      |                       |
| I        | 0.76 (0.55-1.05)     | 0.096                 | 0.81 (0.4859-1.1416) | 0.2179                |
| II       | Reference            | Reference             |                       |                       |
| III      | 2.42 (2.04-2.86)     | < 0.001<sup>c</sup>   | 1.14 (0.9361-1.3457) | 0.2071                |
| IV       | 2.38 (1.82-3.10)     | < 0.001<sup>c</sup>   | 1.15 (0.8608-1.4452) | 0.3396                |
| Tumor size |                 |                       |                      |                       |
| ≤ 30 mm  | Reference            | Reference             |                       |                       |
| 30-50 mm | 1.98 (1.59-2.46)     | < 0.001<sup>c</sup>   | 1.00 (0.7592-1.2460) | 0.9832                |
| 50-65 mm | 2.10 (1.64-2.69)     | < 0.001<sup>c</sup>   | 1.11 (0.8352-1.3831) | 0.4585                |
> 65 mm & 2.67 (2.14-3.34) & < 0.001\(^{c}\) & 1.11 (0.8608-1.4452) & 0.4209 \\
LODDS & ≤ -0.6504 & Reference & Reference & \\
> -0.6504 & 4 (3.44-4.65) & < 0.001\(^{c}\) & 2.20 (1.9068-2.5030) & < 0.001\(^{c}\) \\
LNR & ≤ 0.1795 & Reference & - & - \\
> 0.1795 & 3.97 (3.41-4.62) & < 0.001\(^{c}\) & - & - \\

\(^{a}\)P < 0.05.  
\(^{b}\)P < 0.01.  
\(^{c}\)P < 0.001.

CEA: Carcinoembryonic antigen; LODDS: Log odds of positive lymph nodes; LNR: Lymph node ratio.

### Table 3 Prognostic accuracy of N stage, log odds of positive lymph nodes, and lymph node ratio

|        | C-index | AIC       | AUC 1-yr | AUC 3-yr | AUC 5-yr |
|--------|---------|-----------|----------|----------|----------|
| N stage| 0.6712  | 9906.03   | 0.695    | 0.702    | 0.696    |
| LODDS  | 0.6497  | 9887.95   | 0.690    | 0.673    | 0.660    |
| LNR    | 0.6494  | 9890.20   | 0.689    | 0.673    | 0.659    |

C-index: Harrell concordance index; AIC: Akaike information criterion; AUC: Area under the curve; LODDS: Log odds of positive lymph nodes; LNR: Lymph node ratio.

Although both LODDS and LNR reduced stage migration to some extent, they also have limitations. The classification criteria of both systems did not reach a consensus, and the cutoff value of the LODDS system influenced its performance in predicting the prognosis of patients with CRC after surgery. When analyzed as a continuous variable, the LODDS staging system performed better than the LNR system and the N stage system and it was not influenced by the number of resected LNs. However, when analyzed as a categorical variable, the LNR has more clinical value\[22\]. Another limitation of both systems was that they neglect the TD condition. It is clear that positive TD status is an independent risk factor for a poor prognosis of CRC without metastatic LNs\[26\]. Many studies have shown that patients with TD have a worse...
prognosis or a shorter survival time than those with negative TD\cite{27,28}.

In our study, two opposite results were obtained regarding the predictive accuracy of N stage and LODDS. We found that the predictive accuracy of N stage was better than that of LODDS through the C-index and AUC, which may be caused by the neglect of TD and the choice of the cutoff value in the LODDS system. However, the opposite result was determined through AIC, which may be due to the stage migration of the N stage. The LODDS system is better than the LNR system. Through multivariate analysis, both the LODDS system and TD were independent risk factors for the prognosis of CRC after tumor resection. Compared with the N staging system, the LODDS system can reduce stage migration and discriminate patients in the N0 stage with different numbers of RLNs. Compared with the model with N stage alone, our final nomogram exhibited a better sensitivity and specificity through the likelihood ratio test.

There are still limitations that must be considered regarding our nomogram. As mentioned above, the cutoff value of LODDS influences its predictive performance, so determination of the optimal classification criteria has the potential to improve the predictive efficacy of the nomogram. In our nomogram, LODDS and TD were two separate independent risk factors that influenced the prognosis of CRC patients after surgery. If we can combine the two factors suitably, the nomogram will be simplified further. Additional researches should also be performed to improve the nomogram because there are other factors that affect the OS of postoperative CRC patients. In addition, further prognosis stratification analysis with larger sample sizes is warranted.

**CONCLUSION**

From our research, we found that LODDS performs better than LNR in evaluating the LN metastasis condition of CRC. Age at diagnosis, race, TD, T stage, LODDS, M stage and CEA level are independent prognostic factors of OS in CRC. We built a nomogram
incorporating the above risk factors to predict 1-year, 3-year, and 5-year OS among CRC patients after surgery, which had good predictive accuracy. Our final nomogram performs better than the nomogram with N stage. In clinical practice, the indicators in the nomogram are easy to acquire, and the nomogram can predict OS intuitively. It has guiding significance for doctors making decisions about individualized adjuvant therapeutic schemes for patients before or after surgery.

**ARTICLE HIGHLIGHTS**

**Research background**
A nomogram is an effective tool to predict patient outcomes intuitively. Lymph node (LN) metastasis and tumor deposit (TD) conditions affect the prognosis of patients with colorectal cancer (CRC) after surgery markedly. At present, establishing an effective tool to predict the overall survival (OS) of CRC patients after surgery is necessary.

**Research motivation**
To establish a predictive model to assess the prognosis of CRC patients after surgery.
**Research objectives**

To screen out the suitable risk factors that can affect the OS of CRC patients after surgery and establish a nomogram with these factors.

**Research methods**

A total of 3139 patients diagnosed with CRC after surgery from the Surveillance, Epidemiology, and End Results program were divided into a training set ($n = 2029$) and a validation set ($n = 1047$) randomly. The Harrell concordance index (C-index), Akaike information criterion, and area under the curve (AUC) were used to assess the predictive efficacy of the N stage from the American Joint Committee Cancer tumor-node-metastasis classification, LN ratio, and log odds of positive lymph nodes (LODDS). Construction of the nomogram was based on the risk factors screened out through univariate and multivariate analyses. The C-index, receiver operating characteristic (ROC) curve, calibration curve, and likelihood ratio test were used to assess the final nomogram.

**Research results**

Seven independent predictive factors, namely, race, age at diagnosis, T stage, M stage, LODDS, TD condition, and serum carcinoembryonic antigen level, were included in the nomogram. The C-index of the nomogram for OS prediction was 0.8002 (95%CI: 0.7839-0.8165) in the training set and 0.7864 (95%CI: 0.7604-0.8124) in the validation set. The AUC values of the ROC curve predicting the 1-, 3-, and 5-year OS were 0.846, 0.841, and 0.825, respectively, in the training set and 0.823, 0.817, and 0.835, respectively, in the validation test. The final nomogram showed better sensitivity and specificity than the nomogram with N stage alone for evaluating LN metastasis in both the training set (-4668.0 vs -4688.3, $P < 0.001$) and the validation set (-1919.5 vs -1919.8, $P < 0.001$) through the likelihood ratio test.

**Research conclusions**

The nomogram incorporating LODDS, TD, and other risk factors shows a great predictive accuracy and better sensitivity and specificity and represents a potential tool for therapeutic decision-making.

**Research perspectives**

Suitable combination of LODDS and TD is necessary to simplify the nomogram. Larger sample size studies are required to include more potential risk factors, improve the nomogram, and stratify the prognosis further.

**ACKNOWLEDGEMENTS**

The authors gratefully acknowledge the efforts of the SEER program to create the SEER database.

**REFERENCES**

1. Siegel RL, Miller KD, Goding Sauer A, Fedewa SA, Butterly LF, Anderson JC, Cereck A, Smith RA, Jemal A. Colorectal cancer statistics, 2020. *CA Cancer J Clin* 2020; 70: 145-164 [PMID: 32133645 DOI: 10.3322/caac.21601]
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018; 68: 394-424 [PMID: 30207593 DOI: 10.3322/caac.21492]
3. Amin MB, Edge S, Greene F, Byrd DR, Brookland RK, Washington MK, Gershenwald JE, Compton CC, Hess KR, Sullivan DC, Jessup JM, Brierley JD, Gaspar LE, Schilsky RL, Balch CM, Winchester DP, Asare EA, Madera M, Gress DM, Meyer LR. AJCC Cancer Staging Manual. 8th ed. New York: Springer International Publishing, 2017: 251-274 [DOI: 10.1007/978-3-319-40618-3]
4. Wong SL. Lymph node counts and survival rates after resection for colon and rectal cancer. *Gastrointest Cancer Res* 2009; 3: S33-S35 [PMID: 19461921]
5. Berger AC, Sigurdson ER, LeVoyer T, Hanlon A, Mayer RJ, Macdonald JS, Catalano PJ, Haller DG. Colon cancer survival is associated with decreasing ratio of metastatic to examined lymph nodes. *J Clin Oncol* 2005; 23: 8706-8712 [PMID: 16314630 DOI: 10.1200/jco.2005.02.8852]
6. Shen F, Cui J, Cai K, Pan H, Bu H, Yu F. Prognostic accuracy of different lymph node staging systems in rectal adenocarcinoma with or without preoperative radiation therapy. *Jpn J Clin Oncol*
Li BW et al. Prognostic nomogram for postoperative CRC patients

2018; 48: 625-632 [PMID: 2978392 DOI: 10.1093/jcoco/hby070]

7 Tamura M, Matsumoto I, Saito D, Yoshida S, Takata M, Takemura H. Lymph node ratio as a prognostic factor in patients with pathological N2 non-small cell lung cancer. World J Surg Oncol 2016; 14: 295 [PMID: 27884195 DOI: 10.1186/s12957-016-1048-5]

8 Smith DD, Nelson RA, Schwarz RE. A comparison of five competing lymph node staging schemes in a cohort of resectable gastric cancer patients. Ann Surg Oncol 2014; 21: 875-882 [PMID: 24242677 DOI: 10.1245/s10434-013-3356-0]

9 Persiani R, Cananzi FC, Biondi A, Paliani G, Tufo A, Ferrara F, Vigorita V, D’Ugo D. Log odds of positive lymph nodes in colon cancer: a meaningful ratio-based lymph node classification system. World J Surg 2012; 36: 667-674 [PMID: 22270984 DOI: 10.1007/s00268-011-1415-x]

10 Rencuzogullari A, Benlice C, Valente M, Abbas MA, Remzi FH, Gorgun E. Predictors of Anastomotic Leak in Elderly Patients After Colectomy: Nomogram-Based Assessment From the American College of Surgeons National Surgical Quality Program Procedure-Targeted Cohort. Dis Colon Rectum 2017; 60: 527-536 [PMID: 28338453 DOI: 10.1097/DCR.0000000000000789]

11 Xu Z, Jing J, Ma G. Development and validation of prognostic nomogram based on log odds of positive lymph nodes for patients with gastric signet ring cell carcinoma. Chin J Cancer Res 2020; 32: 778-793 [PMID: 33474000 DOI: 10.1111/jccc.13626]

12 Le Voyer TE, Sigurdson ER, Hanlon AL, Mayer RJ, Macdonald JS, Catalano PJ, Haller DG. Colon cancer survival is associated with increasing number of lymph nodes analyzed: a secondary survey of intergroup trial INT-0089. J Clin Oncol 2003; 21: 2912-2919 [PMID: 12885809 DOI: 10.1200/jco.2003.05.062]

13 Han I, Mo S, Xiang W, Li Q, Wang R, Xu Y, Dai W, Cai G. Comparison of four lymph node staging systems for predicting prognosis for stage IV rectum cancer. Ann Transl Med 2020; 8: 111 [PMID: 32175404 DOI: 10.21037/atm.2019.12.90]

14 J Tsai HL, Lu CY, Hsieh JS, Wu DC, Jhun CM, Chai CY, Chu KS, Chan HM, Wang JY. The prognostic significance of total lymph node harvest in patients with T2-N0M0 colorectal cancer. J Gastrointest Surg 2007; 11: 660-665 [PMID: 17468927 DOI: 10.1007/s11605-007-0119-x]

15 Swansson RS, Compton CC, Stewart AK, Bland KL. The prognostic tier of T3N0 colorectal cancer is dependent on the number of lymph nodes examined. Ann Surg Oncol 2003; 10: 65-71 [PMID: 12513963 DOI: 10.1245/aso.2003.03.058]

16 Wang J, Hassett JM, Dayton MT, Kulaylat MN. Lymph node ratio: role in the staging of node-positive colon cancer. Ann Surg Oncol 2008; 15: 1600-1608 [PMID: 18327530 DOI: 10.1245/s10434-007-9716-x]

17 Huh JW, Kim YJ, Kim HR. Ratio of metastatic to resected lymph nodes as a prognostic factor in node-positive colorectal cancer. Ann Surg Oncol 2010; 17: 2640-2646 [PMID: 20300967 DOI: 10.1245/s10434-010-1015-2]

18 Qiu HB, Zhang LY, Li YF, Zhou ZW, Keshari RP, Xu RH. Ratio of metastatic to resected lymph nodes enhances to predict survival in patients with stage III colorectal cancer. Ann Surg Oncol 2011; 18: 1568-1574 [PMID: 21207155 DOI: 10.1245/s10434-010-1520-7]

19 Arslan NC, Sokmen S, Canda AE, Terzi C, Sarisoglu S. The prognostic impact of the log odds of positive lymph nodes in colon cancer. Colorectal Dis 2014; 16: O386-O392 [PMID: 24980876 DOI: 10.1111/codi.12702]

20 Junginger T, Goener U, Lollert A, Hollemann D, Berres M, Blettner M. The prognostic value of lymph node ratio and updated TNM classification in rectal cancer patients with adequate vs inadequate lymph node dissection. Tech Coloproctol 2014; 18: 805-811 [PMID: 24643761 DOI: 10.1007/s10151-014-1316-x]

21 Park IJ, Yu CS, Lim SB, Yoon YS, Kim CW, Kim TW, Kim JH, Kim JC. Ratio of metastatic lymph nodes is more important for rectal cancer patients treated with preoperative chemoradiotherapy. World J Gastroenterol 2015; 21: 3274-3281 [PMID: 25805934 DOI: 10.3748/wjg.v21.i11.3274]

22 Pei JP, Zhang CD, Fan YC, Dai DQ. Comparison of Different Lymph Node Staging Systems in Patients With Resectable Colorectal Cancer. Front Oncol 2018; 8: 671 [PMID: 30697530 DOI: 10.3389/fonc.2018.00671]

23 Wang C, Yang C, Wang W, Xia B, Li K, Sun F, Hou Y. A Prognostic Nomogram for Cervical Cancer after Surgery from SEER Database. J Cancer 2018; 9: 3923-3928 [PMID: 30401956 DOI: 10.7150/jca.26220]

24 Xiao Z, Shi Z, Hu L, Gao Y, Zhao J, Liu Y, Xu Q, Huang D. A new nomogram from the SEER database for predicting the prognosis of gallbladder cancer patients after surgery. Ann Transl Med 2019; 7: 738 [PMID: 32042754 DOI: 10.21037/atm.2019.11.112]

25 Scarinci A, Di Cesare T, Cavanaglia D, Neri T, Colletti M, Cosenza G, Liverani A. The impact of the log odds of positive lymph nodes (LODDS) in colon and rectal cancer patient stratification: a single-center analysis of 323 patients. Updates Surg Cancer 2018; 70: 23-31 [PMID: 29500795 DOI: 10.1007/s13304-018-0519-3]

26 Zheng K, Zheng N, Xin C, Zhou L, Sun G, Wen R, Zhang H, Yu G, Bai C, Zhang W. The Prognostic Significance of Tumor Deposit Count for Colorectal Cancer Patients after Radical Surgery. Gastroenterol Res Pract 2020; 2020: 2052561 [PMID: 32256564 DOI: 10.1155/2020/2052561]

27 Pricolo VE, Steingrimsson J, McDuffie TJ, McHale JM, McMillen B, Shparber M. Tumor Deposits in Stage III Colon Cancer: Correlation With Other Histopathologic Variables, Prognostic Value, and Risk Stratification-Time to Consider “N2c”. Am J Clin Oncol 2020; 43: 133-138 [PMID: 31764018 DOI: 10.1097/COC.000000000000645]
Liu F, Zhao J, Li C, Wu Y, Song W, Guo T, Chen S, Cai S, Huang D, Xu Y. The unique prognostic characteristics of tumor deposits in colorectal cancer patients. *Ann Transl Med* 2019; 7: 769 [PMID: 32042785 DOI: 10.21037/atm.2019.11.69]
