**Name of Journal:** *World Journal of Clinical Cases*

**Manuscript NO:** 60753

**Manuscript Type:** ORIGINAL ARTICLE

***Retrospective Cohort Study***

**Critical prognostic value of the log odds of negative lymph nodes/tumor size in rectal cancer patients**

Xie JB *et al*. NLNs/tumor size in rectal cancer

Jie-Bin Xie, Yue-Shan Pang, Xun Li, Xiao-Ting Wu

**Jie-Bin Xie, Xiao-Ting Wu,** Department of Gastrointestinal Surgery, West China Hospital, Sichuan University, Chengdu 610041, Sichuan Province, China

**Jie-Bin Xie, Xun Li,** Department of Gastrointestinal Surgery, Affiliated Hospital of North Sichuan Medical College China, Nanchong 637200, Sichuan Province, China

**Yue-Shan Pang,** Department of Geriatrics, The Second Clinical Medical College of North Sichuan Medical College, Nanchong Central Hospital, Nanchong 637200, Sichuan Province, China

**Author contributions:** Xie JB performed design, data collection, analysis, and drafting of manuscript; Pang YS performed design, data collection, and analysis; Li X performed design, data collection, and review; Wu XT performed design, analysis, review and editing; all authors read and approved the final version of the manuscript.

**Supported by** Cooperative Fund of Nanchong Government and North Sichuan Medical College, No. 18SXHZ0357.

**Corresponding author: Xiao-Ting Wu, MD, PhD, Chief Doctor,** Department of Gastrointestinal Surgery, West China Hospital, Sichuan University, No. 37 Guoxue Alley, Wuhou District, Chengdu 610041, Sichuan Province, China. wxt1@medmail.com.cn

**Received:** November 12, 2020

**Revised:** January 2, 2021

**Accepted:** March 10, 2021

**Published online:**

**Abstract**

BACKGROUND

The number of negative lymph nodes (NLNs) and tumor size are associated with prognosis in rectal cancer patients undergoing surgical resection. However, little is known about the prognostic significance of the NLN count after adjusting for tumor size.

AIM

To assess the prognostic impact of the log odds of NLN/tumor size (LONS) in rectal cancer patients.

METHODS

Data of patients with stage I–III rectal cancer were extracted from the Surveillance, Epidemiology, and End Results Program database. These patients were randomly divided into a training cohort and a validation cohort. Univariate and multivariate Cox regression analyses were used to determine the prognostic value of the LONS. The optimal cutoff values of LONS were calculated using the "X-tile" program. Stratified analysis of the effect of LONS on cancer-specific survival (CSS) and overall survival (OS) were performed. The Kaplan-Meier method with the log-rank test was used to plot the survival curve and compare the survival data among the different groups.

RESULTS

In all, 41080 patients who met the inclusion criteria were randomly divided into a training cohort (*n* = 28775, 70%) and a validation cohort (*n* = 12325, 30%). Univariate and multivariate analyses identified the continuous variable LONS as an independent prognostic factor for CSS [training cohort: Hazard ratio (HR) = 0.47, 95% confidence interval (CI): 0.44–0.51, *P* < 0.001; validation cohort: HR = 0.46, 95%CI: 0.41-0.52, *P* < 0.001] and OS (training cohort: HR = 0.53, 95%CI: 0.49-0.56, *P* < 0.001; validation cohort: HR = 0.52, 95%CI: 0.42-0.52, *P* < 0.001). The X-tile program indicated that the difference in CSS was the most significant for LONS of -0.8, and the cutoff value of -0.4 can further distinguish patients with a better prognosis in the high LONS group. Stratified analysis of the effect of the categorical variable LONS on CSS and OS revealed that LONS was also an independent predictor, independent of pN stage, pT stage, tumor-node-metastasis stage, site, age, sex, the number of examined lymph nodes, race, preoperative radiotherapy and carcinoembryonic antigen level.

CONCLUSION

LONS is associated with improved survival of rectal cancer patients independent of other clinicopathological factors.

**Key Words:** Rectal cancer; Negative lymph nodes; Negative lymph nodes/tumor size; Prognosis; Survival analysis; Surveillance, Epidemiology, and End Results Program

Xie JB, Pang YS, Li X, Wu XT. Critical prognostic value of the log odds of negative lymph nodes/tumor size in rectal cancer patients. *World J Clin Cases* 2021; In press

**Core Tip:** Log odds of negative lymph nodes/tumor size (LONS) was defined as the log of the ratio between negative lymph node count and the tumor size and has been first reported as a survival predictive tool in our study. Univariate and multivariate analyses confirmed LONS as an independent prognostic factor for stage I to III rectal cancer. The X-tile program demonstrated that the optimal cutoff was -0.8. Stratified analysis and Kaplan-Meier curves showed a significant improvement in the 10-year cancer-specific survival and overall survival in the high LONS group independent of clinicopathological factors.

**INTRODUCTION**

Rectal cancer (RC) is a common digestive tract tumor. Currently, it is the world's fourth most deadly cancer with almost 900000 deaths annually. Although Western developed countries exhibit a slightly or steady declining trend, the morbidity and mortality of colorectal cancer (CRC) in developing countries, such as China, are also rapidly increasing[1]. Treatment options for CRC have been developed rapidly in the past 20 years; however, surgery is still the cornerstone of curative intent treatment[2]. Of course, postoperative adjuvant radiotherapy and/or chemotherapy in high-risk patients can improve the long-term survival rates[3]; unfortunately, due to the lack of reliable markers, selecting the optimal therapy for individuals is challenging for clinicians[4]. Presently, the most widely used model for risk stratification in CRC is tumor-node-metastasis (TNM) staging. The American Joint Committee on Cancer (AJCC) recommends postoperative adjuvant chemotherapy for patients with high-risk stage II and stage III; however, heterogeneity in prognosis still exists in patients with the same TNM stage[4].

In recent years, biomarkers, such as the number of examined lymph nodes (LNs)[5], negative LNs (NLNs)[6], tumor size[7] and epigenetic biomarkers[8] have increasingly demonstrated independent prognostic value in RC. Among these biomarkers, except for the examined LNs (ELNs) and NLNs that can reflect the degree of lymph node dissection (LND) in surgery, the others mainly represent tumor characteristics or the patient's state. Unfortunately, most of these biomarkers are expensive and difficult to measure. Therefore, to individualize treatment strategies and for clinical application convenience, more prognostic factors should be identified from daily medical records.

The presence of lymph node metastasis is an important prognostic factor for CRC patients. Some studies have revealed that the more ELNs[8] or NLNs[6] that are removed, the fewer LN micrometastases that are missed, and these biomarkers are independent prognostic factors in gastric cancer[9], RC[6], lung cancer[10], and others. However, the number of lymph nodes that can be retrieved varies with the age and sex of the patient and on the tumor grade or site[11], and the skills of the surgeons[12]. Therefore, the same ELNs and NLNs do not mean the same level of LND.

As one of the tumor burden markers, tumor size is defined as the maximal horizontal tumor diameter and has been confirmed to be a prognostic indicator for some solid tumors, including CRC[7,13-15]. Using tumor size to adjust and improve the prognostic value of tumor markers is a novel and common approach, and has been used for carcinoembryonic antigen (CEA)[16] and prostate specific antigen density[17]. Whether the combination of tumor size and NLNs serves as a novel prognostic marker for RC remains unknown.

In the current study, we considered the value of NLNs/tumor size as non-lognormal distribution data, and zero NLN as no logarithm. Therefore, we defined the log odds of negative lymph nodes/tumor size (LONS) as the log of the ratio between the NLN counts plus one and the tumor size, which reflects the NLNs adjusted by the tumor size, to better represent the degree of LND. Our aim was to investigate the prognostic value of LONS in patients with RC.

**MATERIALS AND METHODS**

***Patients***

The Surveillance, Epidemiology, and End Results Program (SEER) collects and provides information on cancer patients such as demographics, tumor characteristics (histology, grading, and TNM stage), treatment and vital status, covering 27.8% of the population in the United States.

Information on patients who were diagnosed with stage I to III RC and underwent a radical excision were collected from the 18 Researcher data (with additional treatment fields) using SEER\*stat 8.3.8 software, which was submitted in November 2018. According to the histological types, including adenocarcinoma (8140/3), mucinous adenocarcinoma (8480/3), and signet-ring cell carcinoma (8490/3), we identified a total of 93496 primary RC patients diagnosed between 2004 and 2015. Only those patients between 18 and 85 years at diagnosis were enrolled.

The exclusion criteria were as follows: (1) the tumor size was unknown; (2) unclear ELNs and NLNs; (3) the T stage or N stage or M stage was unknown; (4) the survival time was 0 mo; and (5) there was no surgical resection or the case lacked a detailed description of the surgery. Patients who satisfied the inclusion criteria were randomly divided into two cohorts (7:3) for cross-validation: A training cohort and a validation cohort. The training cohort was used to generate an optimal cutoff point, and the validation cohort was used to test the applicability of this cutoff point and the final model. Due to the strict register-based nature of the study, informed consent was waived. Moreover, the study was exempted from Institutional Review Board approval, in view of the SEER’s use of unidentifiable patient information.

***Data collection***

Parameters, including preoperative radiation and chemotherapy status, the exact tumor size, regional nodes examined, regional nodes positive, tissue type, tumor differentiation, tumor deposits, CEA levels, survival status, and demographic characteristics were collected. The TNM status of each patient was re-evaluated according to the 8th edition of the AJCC Cancer Staging Manual based on the tumor size, local extension, and LN involvement recorded in the SEER database. Cancer-specific deaths were treated as events. Causes of death that were unclear or deaths from other causes were treated as censored observations. The latest follow-up date occurred in November 2018.

***Statistical analysis***

In our study, the value of the NLNs/size includes non-lognormal distribution data and the zero NLN had no logarithm. Therefore, the data were logarithmically transformed by NLN counts plus one to obtain approximately normal frequency distributions. Continuous and categorical variables were expressed as medians and totals (percentages), respectively. Univariate Cox proportional hazards regression was performed to identify potential prognostic factors. Variables with a *P* < 0.10 in the univariate analysis were selected to include in the multivariate Cox model. Multivariate analysis using the Cox proportional hazards regression model was used to identify independent risk factors and to calculate each hazard ratio (HR). Stratified analyses of the LONS effect on cancer-specific survival (CSS) and overall survival (OS) rates based on different clinicopathological factors were performed by Cox proportional hazards regression. The optimal cutoff values of the LONS were calculated by the X-tile program in terms of CSS[18]. Based on this cutoff, we divided the patients into two groups: The high LONS group and low LONS group. The Kaplan-Meier method with the log-rank test was used to plot the survival curves and compare the survival data among the different groups. All tests were bilateral, and *P* < 0.05 was considered statistically significant. All analyses were performed using SPSS 22.0. The primary outcome was CSS. The secondary endpoint was OS.

**RESULTS**

***Baseline characteristics***

A total of 93496 patients with primary RC from 2004 to 2015 were identified in the SEER database. Among them, 30795 did not undergo surgical resection or detailed surgical information was lacking, 52 were not histologically confirmed to have adenocarcinoma, and for 10120 patients, their record failed to include the exact tumor size. Of the remaining 52529 patients, 2597 patients were excluded due to no LNs removed or unclear ELNs or PLNs, 6791 patients were not classified as stage I to III, 1734 patients were not between 18 and 85 years of age, and 327 had missing survival time information. After excluding these patients, 41080 patients satisfied the inclusion criteria (Figure 1). The final cases were randomly divided into the training cohort (*n* = 28775, 70%) and the validation cohort (*n* = 12325, 30%).

Among these patients, the median (interquartile range) age, tumor size, ELN count, and NLN count were 62 (53-72) years old, 40 (27-55) mm, 14 (10-20), and 13 (9-18), respectively. In total, 16998 (59.1%) patients were male, 11757 (40.1%) were female. A total of 18661 (64.9%) tumors were located in the rectum, and 10094 (35.1%) were located in the rectosigmoid junction. In addition, 10385 (36.1%) and 17688 (61.5%) patients received preoperative radiotherapy and chemotherapy, respectively. The proportions of N0, N1 and N2 stage were 61.3%, 25.3% and 13.3%, respectively. A similar distribution of features was observed in patients in the validation cohort. Detailed information on the clinical pathological features is shown in Table 1.

***Univariate and multivariate analysis***

Age, sex, location, race, TNM stage, tumor size, differentiation, histology type, preoperative radiotherapy, chemotherapy, preoperative CEA and LONS were identified as potential prognostic factors by univariate analysis for CSS (Table 2) and OS (Supplementary Table 1) in both cohorts. The HR showed an obvious improvement for LONS [0.330, 95% confidence interval (CI): 0.308-0.354], which was the NLNs adjusted by the tumor size, compared with the NLNs (0.975, 95%CI: 0.967-0.974) in CSS. Similar results were confirmed for the OS. Multivariate analysis with Cox regression for CSS (Table 3) and OS (Supplementary Table 2) in both cohorts were performed. All of these factors with the exception of tumor size were independent prognostic factors for RC. Similar results were observed in patients from the validation cohort.

***Optimal cutoff values of LONS and Kaplan-Meier curves***

The "X-tile" program was then used to obtain the optimal cutoff values of LONS, and the highest chi-square log-rank value of 613.1 was produced in terms of CSS when applying -0.8 as the cutoff value of LONS. This cutoff divided the LONS into the high LONS group and the low LONS group. Kaplan-Meier curves revealed a significant improvement in the 10-year CSS (Figure 2A) and OS (Figure 2B) in the high LONS group compared with the low LONS group in the training cohort. The better outcomes of those with a high LONS was confirmed in the validation cohort as these patients exhibited a better 10-year CSS (Figure 3A) and OS (Figure 3B).

Moreover, the cutoff values of -0.9 and -0.4 also had discriminatory power for different prognoses with a maximum of the chi-square log-rank value of 845.3. This cutoff divided the LONS into three subsets, the 10-year CCS (Figure 2C) and OS (Figure 2D) differed for the high, middle and low group. The statistical results of the validation cohort were consistent with those of the training cohort, and the 10-year CCS (Figure 3C) and OS (Figure 3D) were different for the high, middle and low groups.

***Stratified analysis of the LONS effect on CSS and OS rates based on the different clinicopathological factors***

To confirm the independent prognostic effect of LONS on the different clinicopathological factors, the role of LONS on CSS and OS were further studied by stratified analysis with a multivariate Cox proportional hazards model. Among the subsets of the pN stage, pT stage, and TNM stage, sex, age, location, preoperative radiotherapy, CEA, ELNs, differentiation and chemotherapy but not pT4b, N1c and G1, LONS were significantly associated with CSS as categorical variables in the training cohort, and the HR increased strikingly with the increment of ELNs and degree of differentiation (Table 4, reference was the high LONS group). For OS, only in the subgroup of N1c (Supplementary Table 3), LONS was not an independent prognostic factor. Similar results were obtained in the validation group cohort. However, Kaplan-Meier curves of pT, pN and pTNM, including pT4b, pN1c and G1, revealed that a higher LONS was significantly correlated with improved CSS (Figures 4 and 5) and OS (Supplementary Figure 1) in each subset. The validation cohort showed similar results (Supplementary Figures 2 and 3).

**DISCUSSION**

In the present study, the large population-based SEER database was first used to quantify the relative degree of LND and to evaluate the prognostic value of LONS for stage I to III RC. Our results demonstrated that high LONS was associated with improved survival of RC patients. Subgroup analysis confirmed that LONS was an independent prognostic factor independent of the clinicopathological factors.

Previous studies suggested that ELN[11] and NLN[6] were correlated with survival of RC independent of lymph node metastasis and tumoral molecular alterations. In addition, the greater the number of negative lymph nodes, the better the level of LND was. However, given that the number of lymph nodes that can be retrieved varies by age and sex of the patient, the tumor grade or site, and the skills of the surgeons and that NLN lacks information about the biological characteristics of the tumors, the value of NLN in survival prediction remains underappreciated. Thus, NLN exhibits limitations in clinical applications, and consensus about its utility is lacking in the literature.

As one of the major tumor characteristics, tumor size is not only related to the prognosis[7], but also closely related to its biological characteristics[19]. Therefore, we decided to use LONS, which is defined as the log of the ratio between the number of negative nodes plus one and the tumor size and serves as the index of the average number of NLNs *per* tumor unit. The higher the value, the more negative lymph nodes that were obtained. Conversely, a lower LONS value indicates fewer NLNs *per* tumor unit. Therefore, LONS can be used to compare the relative level of LND among different patients.

In our study, univariate analyses indicated that the HR of NLNs was 0.975 (95%CI: 0.967-0.974), and the HR of NLNs adjusted by tumor size was 0.330 (95%CI: 0.308-0.354). Our results revealed that using tumor size to adjust NLNs may significantly improve the prognostic value of NLNs. We confirmed the independent prognostic value of LONS among all patients with stage I–III RC by multivariate Cox analysis and in a validation cohort. A LONS cutoff value of -0.8 divided patients into the high LONS group and the low LONS group. Kaplan-Meier curves showed that patients with a high LONS (over -0.8) had a significantly better 10-year CSS and OS. A more encouraging result was that a cutoff value of -0.4 can further distinguish patients with a better prognosis in the high LONS group. Thus, postoperative LONS can be applied as a novel prognostic marker for patients with stage I–III RC.

Notably, to confirm the independent prognostic effect of LONS, we stratified the analysis of the LONS effect on the CSS and OS rates based on different clinicopathological factors, including CEA level, preoperative radiotherapy and chemotherapy, age, sex, location, differentiation, histology type, the number of ELNs, TNM stage, pT stage and pN stage. Excitingly, this study demonstrated that LONS was associated with improved survival of RC patients independent of their clinicopathological factors, and the HR increased strikingly with the increment of ELNs and degree of differentiation. Even N0 stage patients with a poor prognosis can be distinguished from these patients, who are generally considered to have a good outcome using LONS. Although there was no significant difference in T4b, N1c or the G1 subsets in the multivariate Cox analysis, survival analysis still demonstrated that patients with high LONS exhibited a better survival benefit. These results suggested that LONS is an independent prognostic factor that is independent of tumor characteristics and can be used as an index of relative LND. This marker can be simply calculated from the postoperative pathological report, at no extra cost. With widespread clinical use, additional improvements in the accuracy of predicting 10-year outcomes will benefit more patients. Special attention should be given to lower LONS patients with stage I and no high-risk stage IIa tumors that may be underestimated as stage I or II due to insufficient LND. These patients should undergo early clinical interventions. For stage III patients with a lower LONS, systematic and intensive treatment should be given more actively to achieve the same therapeutic effect.

However, we admit that there were some inherent limitations in our study. First, using tumor size to represent tumor characteristics is not precise because the histological type, differentiation degree and genotyping of RC are also important biological characteristics that are not included in our adjustment factors. In addition, the tumor size cannot represent the actual tumor volume or tumor load[20]; however, tumor size is closely related to tumor grade[15], tissue differentiation[21] and other biological characteristics[22]. For clinical application, tumor size represents a convenient and quick method that can be applied to roughly estimate the tumor volume[16]. Second, beyond tumor size, NLNs are also related to the above factors, that are not included in the adjustment scope of this study and may affect the judgment of the degree of LND. Third, due to the retrospective nature of the SEER, selection bias could not be completely avoided. Furthermore, our results were not validated in an external database. Whether LONS is truly associated with CSS and OS requires further study.

**CONCLUSION**

In summary, patients with high LONS have a better outcome than those with low LONS. LONS is an independent prognostic factor that is independent of clinicopathological features and can be used as a relative index for the degree of LND. LONS can be used as a novel marker for risk stratification and therapeutic decision-making in RC patients after surgery.

**ARTICLE HIGHLIGHTS**

***Research background***

Rectal cancer (RC) is the world's fourth most deadly cancer with almost 900000 deaths annually. Therapy options for RC have been developed rapidly in the past decade, postoperative adjuvant radiotherapy and/or chemotherapy in high-risk patients can improve their long-term survival rates; however, the clinical outcomes among RC patients with the same tumor-node-metastasis (TNM) stage might be completely different. Unfortunately, due to the lack of reliable markers, selecting the optimal therapy for individuals is challenging for clinicians.

***Research motivation***

The degree of lymph node dissection (LND) is closely related to the prognosis of RC; however, so far, there is no objective and effective evaluation index for LND. Previous studies have suggested that examined lymph nodes (ELNs), negative lymph nodes (NLNs) and size were closely related to the prognosis of RC. Other studies have added another factor such as tumor size to improve the prognostic value of biomarkers, for example preoperative serum carcinoembryonic antigen (CEA) and prostate-specific antigen density. Therefore, we defined a novel prognostic score, the log odds of NLN/tumor size (LONS), as the log of the ratio between the NLN counts plus one and the tumor size, which reflects the NLNs adjusted by the tumor size, to better represent the degree of LND.

***Research objectives***

Our aim was to assess a potentially novel prognostic score to stratify risks for RC patients. At the same time, we also aimed to investigate whether LONS can distinguish different pathological stages and clinical features, to better guide the treatment strategies and follow-up plan.

***Research methods***

The data of stage I–III RC patients were extracted from the Surveillance, Epidemiology, and End Results Program (SEER) database from 2004 to 2017. Univariate and multivariate Cox regression analyses were applied to determine the prognostic impact of the LONS. The optimal cutoff values of the LONS were calculated using the "X-tile" program. Stratified analysis of the LONS effect on cancer-specific survival (CSS) and overall survival (OS) were performed. The Kaplan-Meier method with the log-rank test was used to plot the survival curve and compare the survival data among the different groups.

***Research results***

In all, 41080 patients were finally included in the study and randomly divided into a training cohort (*n* = 28775, 70%) and a validation cohort (*n* = 12325, 30%). Univariate and multivariate analyses identified the continuous variable LONS as an independent prognostic factor for CSS and OS. The X-tile program indicated that the difference in CSS was the most significant for LONS of -0.8，and the cutoff value of -0.4 can further distinguish patients with a better prognosis in the high LONS group. Stratified analysis of the effect of the categorical variable LONS on CSS and OS revealed that LONS was also an independent predictor independent of pN stage, pT stage, TNM stage, site, age, sex, the number of ELNs, race, preoperative radiotherapy and CEA level.

***Research conclusions***

Patients with high LONS have a better outcome than those with low LONS. LONS is an independent prognostic factor that is independent of clinicopathological features and can serve as a relative index for the degree of LND. LONS can be used as a novel marker for risk stratification and therapeutic decision-making in RC patients after surgery.

***Research perspectives***

Due to the retrospective nature of the SEER database, we cannot obtain high-level clinical evidence, but our study provides a novel approach for the evaluation of LND and suggests a potentially novel prognostic score to stratify risks for RC patients at the same stage.

**REFERENCES**

1 **Arnold M**, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. *Gut* 2017; **66**: 683-691 [PMID: 26818619 DOI: 10.1136/gutjnl-2015-310912]

2 **Dekker E**, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB. Colorectal cancer. *Lancet* 2019; **394**: 1467-1480 [PMID: 31631858 DOI: 10.1016/S0140-6736(19)32319-0]

3 **Puccini A**, Lenz HJ. Colorectal cancer in 2017: Practice-changing updates in the adjuvant and metastatic setting. *Nat Rev Clin Oncol* 2018; **15**: 77-78 [PMID: 29182161 DOI: 10.1038/nrclinonc.2017.185]

4 **Dienstmann R**, Salazar R, Tabernero J. Personalizing colon cancer adjuvant therapy: selecting optimal treatments for individual patients. *J Clin Oncol* 2015; **33**: 1787-1796 [PMID: 25918287 DOI: 10.1200/JCO.2014.60.0213]

5 **Lei P**, Ruan Y, Liu J, Zhang Q, Tang X, Wu J. Prognostic impact of the number of examined lymph nodes in stage II colorectal adenocarcinoma: A retrospective study. *Gastroenterol Res Pract* 2020; **2020**: 8065972 [PMID: 32676106 DOI: 10.1155/2020/8065972]

6 **Ogino S**, Nosho K, Irahara N, Shima K, Baba Y, Kirkner GJ, Mino-Kenudson M, Giovannucci EL, Meyerhardt JA, Fuchs CS. Negative lymph node count is associated with survival of colorectal cancer patients, independent of tumoral molecular alterations and lymphocytic reaction. *Am J Gastroenterol* 2010; **105**: 420-433 [PMID: 19809407 DOI: 10.1038/ajg.2009.578]

7 **Li X**, An B, Ma J, He B, Qi J, Wang W, Qin C, Zhao Q. Prognostic value of the tumor size in resectable colorectal cancer with different primary locations: A retrospective study with the propensity score matching. *J Cancer* 2019; **10**: 313-322 [PMID: 30719125 DOI: 10.7150/jca.26882]

8 **Jung G**, Hernández-Illán E, Moreira L, Balaguer F, Goel A. Epigenetics of colorectal cancer: biomarker and therapeutic potential. *Nat Rev Gastroenterol Hepatol* 2020; **17**: 111-130 [PMID: 31900466 DOI: 10.1038/s41575-019-0230-y]

9 **Hayashi S**, Kanda M, Ito S, Mochizuki Y, Teramoto H, Ishigure K, Murai T, Asada T, Ishiyama A, Matsushita H, Tanaka C, Kobayashi D, Fujiwara M, Murotani K, Kodera Y. Number of retrieved lymph nodes is an independent prognostic factor after total gastrectomy for patients with stage III gastric cancer: propensity score matching analysis of a multi-institution dataset. *Gastric Cancer* 2019; **22**: 853-863 [PMID: 30483985 DOI: 10.1007/s10120-018-0902-2]

10 **Liang W**, He J, Shen Y, Shen J, He Q, Zhang J, Jiang G, Wang Q, Liu L, Gao S, Liu D, Wang Z, Zhu Z, Ng CS, Liu CC, Horsleben Petersen R, Rocco G, D'Amico T, Brunelli A, Chen H, Zhi X, Liu B, Yang Y, Chen W, Zhou Q, He J. Impact of examined lymph node count on precise staging and long-term survival of resected non-small-cell lung cancer: A population study of the US SEER database and a Chinese multi-institutional registry. *J Clin Oncol* 2017; **35**: 1162-1170 [PMID: 28029318 DOI: 10.1200/JCO.2016.67.5140]

11 **Sarli L**, Bader G, Iusco D, Salvemini C, Mauro DD, Mazzeo A, Regina G, Roncoroni L. Number of lymph nodes examined and prognosis of TNM stage II colorectal cancer. *Eur J Cancer* 2005; **41**: 272-279 [PMID: 15661553 DOI: 10.1016/j.ejca.2004.10.010]

12 **Becerra AZ**, Aquina CT, Berho M, Boscoe FP, Schymura MJ, Noyes K, Monson JR, Fleming FJ. Surgeon-, pathologist-, and hospital-level variation in suboptimal lymph node examination after colectomy: Compartmentalizing quality improvement strategies. *Surgery* 2017; **161**: 1299-1306 [PMID: 28088321 DOI: 10.1016/j.surg.2016.11.029]

13 **Zhang W**, Jin K, Wang F, Zhangyuan G, Yu W, Liu Y, Zhang H, Zhang P, Sun B. Differences in the prognostic value of tumor size on hepatocellular cancer-specific survival stratified by gender in a SEER population-based study. *United European Gastroenterol J* 2019; **7**: 933-941 [PMID: 31428418 DOI: 10.1177/2050640619845602]

14 **Al Natour RH**, Saund MS, Sanchez VM, Whang EE, Sharma AM, Huang Q, Boosalis VA, Gold JS. Tumor size and depth predict rate of lymph node metastasis in colon carcinoids and can be used to select patients for endoscopic resection. *J Gastrointest Surg* 2012; **16**: 595-602 [PMID: 22143420 DOI: 10.1007/s11605-011-1786-1]

15 **Spolverato G**, Ejaz A, Kim Y, Sotiropoulos GC, Pau A, Alexandrescu S, Marques H, Pulitano C, Barroso E, Clary BM, Aldrighetti L, Bauer TW, Walters DM, Groeschl R, Gamblin TC, Marsh W, Nguyen KT, Turley R, Popescu I, Hubert C, Meyer S, Gigot JF, Mentha G, Pawlik TM. Tumor size predicts vascular invasion and histologic grade among patients undergoing resection of intrahepatic cholangiocarcinoma. *J Gastrointest Surg* 2014; **18**: 1284-1291 [PMID: 24841438 DOI: 10.1007/s11605-014-2533-1]

16 **Cai D**, Huang ZH, Yu HC, Wang XL, Bai LL, Tang GN, Peng SY, Li YJ, Huang MJ, Cao GW, Wang JP, Luo YX. Prognostic value of preoperative carcinoembryonic antigen/tumor size in rectal cancer. *World J Gastroenterol* 2019; **25**: 4945-4958 [PMID: 31543685 DOI: 10.3748/wjg.v25.i33.4945]

17 **Eichelberger LE**, Koch MO, Eble JN, Ulbright TM, Juliar BE, Cheng L. Maximum tumor diameter is an independent predictor of prostate-specific antigen recurrence in prostate cancer. *Mod Pathol* 2005; **18**: 886-890 [PMID: 15803186 DOI: 10.1038/modpathol.3800405]

18 **Camp RL**, Dolled-Filhart M, Rimm DL. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. *Clin Cancer Res* 2004; **10**: 7252-7259 [PMID: 15534099 DOI: 10.1158/1078-0432.CCR-04-0713]

19 **Bermúdez M**, Aguilar-Medina M, Lizárraga-Verdugo E, Avendaño-Félix M, Silva-Benítez E, López-Camarillo C, Ramos-Payán R. LncRNAs as regulators of autophagy and drug resistance in colorectal cancer. *Front Oncol* 2019; **9**: 1008 [PMID: 31632922 DOI: 10.3389/fonc.2019.01008]

20 **Tayyab M**, Razack A, Sharma A, Gunn J, Hartley JE. Correlation of rectal tumor volumes with oncological outcomes for low rectal cancers: does tumor size matter? *Surg Today* 2015; **45**: 826-833 [PMID: 25377268 DOI: 10.1007/s00595-014-1068-0]

21 **Bhindi B**, Thompson RH, Lohse CM, Mason RJ, Frank I, Costello BA, Potretzke AM, Hartman RP, Potretzke TA, Boorjian SA, Cheville JC, Leibovich BC. The Probability of Aggressive Versus Indolent Histology Based on Renal Tumor Size: Implications for Surveillance and Treatment. *Eur Urol* 2018; **74**: 489-497 [PMID: 30017400 DOI: 10.1016/j.eururo.2018.06.003]

22 **Kumagai A**, Kondo F, Sano K, Inoue M, Fujii T, Hashimoto M, Watanabe M, Soejima Y, Ishida T, Tokairin T, Saito K, Sasajima Y, Takahashi Y, Uozaki H, Fukusato T. Immunohistochemical study of hepatocyte, cholangiocyte and stem cell markers of hepatocellular carcinoma: the second report: relationship with tumor size and cell differentiation. *J Hepatobiliary Pancreat Sci* 2016; **23**: 414-421 [PMID: 27161394 DOI: 10.1002/jhbp.356]

**Footnotes**

**Institutional review board statement:** This study was exempted from Institutional Review Board approval, in view of the SEER’s use of unidentifiable patient information. Due to the strict register-based nature of the study, informed consent was waived. We also received permission to obtain research data from the SEER database, No. 11112-Nov2019.

**Informed consent statement:** Patients were not required to provide informed consent for the study because the analysis used anonymous clinical data.

**Conflict-of-interest statement:** The authors declare that they have no conflict of interest.

**Data sharing statement:** No additional data are available.

**STROBE statement:** The authors have read the STROBE Statement-checklist of items, and the manuscript was prepared and revised according to the STROBE Statement-checklist of items.

**Open-Access:** This article is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/Licenses/by-nc/4.0/

**Manuscript source:** Unsolicited manuscript

**Peer-review started:** November 12, 2020

**First decision:** December 24, 2020

**Article in press:**

**Specialty type:** Medicine, research and experimental

**Country/Territory of origin:** China

**Peer-review report’s scientific quality classification**

Grade A (Excellent): 0

Grade B (Very good): 0

Grade C (Good): C

Grade D (Fair): 0

Grade E (Poor): 0

**P-Reviewer:** Linnebacher M **S-Editor:** Fan JR **L-Editor:** Webster JR **P-Editor:**

**Figure Legends**



**Figure 1 Flowchart of patient selection.** SEER: Surveillance, Epidemiology, and End Results Program; ELNs: Examined lymph nodes; PLNs: Positive lymph nodes.



**Figure 2 Kaplan-Meier survival curves of the cutoff values in the training cohort.** A: Kaplan Meier survival curve of cutoff -0.8 cancer-specific survival in the training cohort; B: Kaplan Meier survival curve of cutoff -0.8 overall survival in the training cohort; C: Kaplan Meier survival curve of cutoff -0.9 and -0.4 cancer-specific survival in the training cohort; D: Kaplan Meier survival curve of cutoff -0.9 and -0.4 overall survival in the training cohort.



**Figure 3 Kaplan-Meier survival curves of cutoff values in the validation cohort.** A: Kaplan Meier survival curve of cutoff -0.8 cancer-specific survival in the validation cohort; B: Kaplan Meier survival curve of cutoff -0.8 overall survival in the validation cohort; C: Kaplan Meier survival curve of cutoff -0.9 and -0.4 cancer-specific survival in the validation cohort; D: Kaplan Meier survival curve of cutoff -0.9 and -0.4 overall survival in the validation cohort.



****

**Figure 4 Kaplan-Meier survival curves of T stage and N stage subgroups for cancer-specific survival in the training cohort.**



**Figure 5 Kaplan-Meier survival curves of tumor-node-metastasis stage subgroup for cancer-specific survival and overall survival in the training cohort.**

**Table 1 Baseline clinicopathological features of patients in the training cohort and validation cohort**

|  |  |  |
| --- | --- | --- |
| **Feature** | **Training cohort (*n* = 28875)** | **Validation cohort (*n* = 12325)** |
| ***n* (%)** | ***n* (%)** |
| Median age (IQR), yr | 62 (53-72) | 62 (53-71) |
| Median tumor size (IQR) mm | 40 (27-55) | 40 (26-55) |
| Median ELN count (IQR)  | 14 (10-20) | 15 (10-20) |
| Median NLN count (IQR) | 13 (9-18) | 13 (9-18) |
| Median LONS (IQR) | -0.44 (-0.66 to -0.24) | -0.44 (-0.65 to -0.22) |
| Sex |  |  |
| Male | 16998 (59.1) | 7262 (58.9) |
| Female | 11757 (40.9) | 5063 (41.1) |
| TNM stage |  |  |
| I | 7280 (25.3) | 3132(25.4) |
| II | 10355 (36.0) | 4662 (36.2) |
| III | 11120 (38.7) | 4731 (38.4) |
| Differentiation  |  |  |
| Poor | 3741 (13.0) | 1606 (13.0) |
| Moderate | 21590 (75.1) | 9169 (74.4) |
| High | 2032 (7.1) | 893 (7.2) |
| Unknown | 1392 (4.8) | 657 (5.3) |
| Preoperative radiotherapy |  |  |
| Yes | 10385 (36.1) | 4419 (35.9) |
| No | 18730 (63.9) | 7906 (64.1) |
| Chemotherapy |  |  |
| Yes | 17688 (61.5) | 7501 (60.9) |
| No | 11067 (38.5) | 4824 (39.1) |
| CEA |  |  |
| Normal | 1056 (38.4) | 4712(38.5) |
| High | 6676 (23.2) | 1736 (16.0) |
| Unknown | 11023 (38.3) | 4924 45.5) |
| N stage |  |  |
| N0 | 17635 (61.3) | 7594 (61.6) |
| N1 | 7288 (25.3) | 3036 (24.9) |
| N2 | 3832 (13.3) | 1668 (13.5) |
| Histology type |  |  |
| Adenocarcinoma | 27233 (94.7) | 11640 (94.4) |
| MAC and SRCC | 1522 (5.3) | 5.6 (13.5) |
| Location |  |  |
| Rectum | 18661 (64.9) | 8044 (65.3) |
| Rectosigmoid junction | 10094 (35.1) | 4281 (34.7) |
| Race |  |  |
| White | 23264 (80.9) | 9962 (81.1) |
| Black | 2357 (8.2) | 1028 (8.3) |
| Others | 3062 (10.6) | 1299 (10.5) |

CEA: Carcinoembryonic antigen; ELNs: Examined lymph nodes; IQR: Interquartile range; LONS: Log odds of negative lymph nodes/tumor size; MAC: Mucinous adenocarcinoma; NLNs: Negative lymph nodes; SRCC: Signet-ring cell carcinoma; TNM: Tumor-node-metastasis.

**Table 2 Univariate analysis of prognostic factors for cancer-specific survival**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Training cohort (*n* = 371)** | **Validation cohort (*n* = 12325)** |
| **HR** | **95%CI** | ***P* value** | **HR** | **95%CI** | ***P* value** |
| Age1 | 1.022 | 1.019-1.024 | < 0.001 | 1.021 | 1.018-1.025 | < 0.001 |
| Tumor size1 | 1.004 | 1.003-1.004 | < 0.001 | 1.003 | 1.002-1.003 | < 0.001 |
| NLNs/tumor size1 | 0.971 | 0.967-0.974 | < 0.001 | 0.970 | 0.964-0.975 | < 0.001 |
| LONS1 | 0.330 | 0.308-0.354 | < 0.001 | 0.331 | 0.298-0.367 | < 0.001 |
| Sex (ref = male) | 0.887 | 0.840-0.937 | < 0.001 | 0.901 | 0.829-0.979 | < 0.001 |
| TNM (ref = stage III) |  |  |  |  |  |  |
| I | 0.226 | 0.207-0.247 | < 0.001 | 0.235 | 0.206-0.269 | < 0.001 |
| II | 0.482 | 0.454-0.512 | < 0.001 | 0.486 | 0.443-0.532 | < 0.001 |
| Differentiation (ref = G3 + G4) |  |  |  |  |  |  |
| G1 | 0.436 | 0.382-0.497 | < 0.001 | 0.424 | 0.348-0.517 | < 0.001 |
| G2 | 0.573 | 0.536-0.613 | < 0.001 | 0.566 | 0.511-0.626 | < 0.001 |
| Unknown | 0.532 | 0.459-0.616 | < 0.001 | 0.515 | 0.415-0.641 | < 0.001 |
| Histology (ref = MAC and SRCC) | 0.537 | 0.488-0.590 | < 0.001 | 0.531 | 0.462-0.612 | < 0.001 |
| Preoperative RT (ref = Yes) | 0.884 | 0.836-0.934 | < 0.001 | 0.938 | 0.861-1.020 | 0.135 |
| Chemotherapy (ref = Yes) | 0.746 | 0.705-0.790 | < 0.001 | 0.776 | 0.712-0.845 | < 0.001 |
| Location (ref = rectum) | 1.134 | 1.072-1.200 | < 0.001 | 1.177 | 1.079-1.284 | < 0.001 |
| Race (ref = others) |  |  |  |  |  |  |
| White | 1.045 | 0.955-1.143 | 0.336 | 1.013 | 0.883-1.162 | 0.853 |
| Black | 1.435 | 1.275-1.616 | < 0.001 | 1.465 | 1.225-1.753 | < 0.001 |
| CEA (ref = High) |  |  |  |  |  |  |
| Normal | 0.546 | 0.510-0.584 | < 0.001 | 0.498 | 0.448-0.553 | < 0.001 |
| Unknown | 0.684 | 0.641-0.729 | < 0.001 | 0.686 | 0.622-0.756 | < 0.001 |

1These variables were treated as continuous data. CI: Confidence interval; CEA: Carcinoembryonic antigen; HR: Hazard ratio; LONS: Log odds of negative lymph nodes/tumor size; MAC: Mucinous adenocarcinoma; NLNs: Negative lymph nodes; RT: Radiotherapy; SRCC: Signet-ring cell carcinoma; TNM: Tumor-node-metastasis.

**Table 3 Multivariate analysis of prognostic factors for cancer-specific survival**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Training cohort (*n* = 28755)** | **Validation cohort (*n* = 12325)** |
| **HR** | **95%CI** | ***P*2 value** | **HR** | **95%CI** | ***P*2 value** |
| Age (ref: ≤ 60 years old) | 1.54 | 1.46-1.63 | < 0.001 | 1.50 | 1.37-1.62 | < 0.001 |
| Tumor size1  | 1.00 | 1.00-1.00 | 0.637 | 1.00 | 1.00-1.00 | 0.987 |
| LONS1  | 0.47 | 0.44-0.51 | < 0.001 | 0.46 | 0.41-0.52 | < 0.001 |
| Sex (ref = male) | 0.90 | 0.85-0.95 | < 0.001 | 0.91 | 0.83-0.99 | 0.021 |
| TNM1 (ref = stage III) |  |  |  |  |  |  |
| I | 0.28 | 0.22-0.26 | < 0.001 | 0.264 | 0.23-0.31 | < 0.001 |
| II | 0.47 | 0.45-0.51 | < 0.001 | 0.50 | 0.46-0.55 | < 0.001 |
| Differentiation (ref = G3 + G4) |  |  |  |  |  |  |
| G1 | 0.63 | 0.56-0.72 | < 0.001 | 0.60 | 0.49-0.73 | < 0.001 |
| G2 | 0.72 | 0.67-0.77 | < 0.001 | 0.71 | 0.62-0.82 | < 0.001 |
| Unknown | 0.63 | 0.54-0.73 | < 0.001 | 0.63 | 0.51-0.79 | < 0.001 |
| Histology (ref = MAC and SRCC) | 0.78 | 0.71-0.86 | < 0.001 | 0.71 | 0.62-0.82 | < 0.001 |
| Preoperative RT (ref = Yes) | 0.90 | 0.85-0.96 | < 0.001 | 0.92 | 0.83-1.03 | 0.147 |
| Chemotherapy (ref = Yes) | 1.42 | 1.32-1.52 | < 0.001 | 1.36 | 1.22-1.51 | < 0.001 |
| Location (ref = rectum) | 0.82 | 0.77-0.87 | < 0.001 | 0.77 | 0.70-0.85 | < 0.001 |
| Race (ref = others) |  |  |  |  |  |  |
| White | 1.10 | 1.00-1.20 | 0.043 | 1.04 | 0.91-1.20 | 0.555 |
| Black | 1.45 | 1.29-1.63 | < 0.001 | 1.49 | 1.25-1.79 | < 0.001 |
| CEA (ref = High) |  |  |  |  |  |  |
| Normal | 0.66 | 0.62-0.71 | < 0.001 | 0.60 | 0.54-0.67 | < 0.001 |
| Unknown | 0.80 | 0.75-0.85 | < 0.001 | 0.77 | 0.7-0.86 | < 0.001 |

1These variables were treated as continuous data; 2*P* values were adjusted for sex, age, race, tumor-node-metastasis stage, location, preoperative radiotherapy, carcinoembryonic antigen, differentiation and chemotherapy as covariates. CI: Confidence interval; CEA: Carcinoembryonic antigen; HR: Hazard ratio; LONS: Log odds of negative lymph nodes/tumor size; MAC: Mucinous adenocarcinoma; SRCC: Signet-ring cell carcinoma; RT: Radiotherapy; TNM: Tumor-node-metastasis.

**Table 4 Multivariate analyses for evaluating log odds of negative lymph nodes/tumor size effect on cancer-specific survival based on different clinicopathological factors1**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Training cohort (*n* = 28755)** | **Validation cohort (*n* = 12325)** |
| **HR** | **95%CI** | ***P*2 value** | **HR** | **95%CI** | ***P*2 value** |
| All | 1.673 | 1.57-1.78 | < 0.001 | 1.715 | 1.56-1.86 | < 0.001 |
| N stage |  |  |  |  |  |  |
| N0 | 1.53 | 1.37-1.71 | < 0.001 | 1.422 | 1.20-1.68 | < 0.001 |
| N1a | 1.49 | 1.24-1.78 | < 0.001 | 1.73 | 1.26-2.36 | 0.001 |
| N1b | 1.42 | 1.21-1.65 | < 0.001 | 1.61 | 1.28-2.04 | < 0.001 |
| N1c | 1.68 | 0.96-2.92 | 0.069 | 2.12 | 0.98-4.60 | 0.057 |
| N2a | 1.64 | 1.40-1.93 | < 0.001 | 1.7 | 1.33-2.18 | < 0.001 |
| N2b | 1.55 | 1.34-1.79 | < 0.001 | 1.37 | 1.10-1.71 | 0.005 |
| T stage |  |  |  |  |  |  |
| T1 | 1.69 | 1.47-1.95 | < 0.001 | 1.96 | 1.19-3.22 | 0.008 |
| T2 | 1.42 | 1.16-1.73 | 0.001 | 1.6 | 1.18-2.16 | 0.003 |
| T3 | 1.63 | 1.56-1.80 | < 0.001 | 1.653 | 1.48-1.85 | < 0.001 |
| T4a | 1.51 | 1.18-1.94 | 0.001 | 1.64 | 1.12-2.40 | 0.011 |
| T4b | 1.22 | 0.97-1.53 | 0.084 | 1.52 | 1.04-2.21 | 0.03 |
| TNM stage |  |  |  |  |  |  |
| I | 1.68 | 1.54-1.87 | < 0.001 | 1.59 | 1.13-2.23 | 0.007 |
| II | 1.55 | 1.37-1.75 | < 0.001 | 1.36 | 1.12-1.65 | 0.002 |
| III | 1.73 | 1.60-1.87 | < 0.001 | 1.79 | 1.60-2.02 | < 0.001 |
| Sex |  |  |  |  |  |  |
| Male | 1.722 | 1.59-1.86 | < 0.001 | 1.715 | 1.52-1.94 | < 0.001 |
| Female | 1.588 | 1.44-1.76 | < 0.001 | 1.715 | 1.47-1.99 | < 0.001 |
| Age, yr |  |  |  |  |  |  |
| ≤ 60 | 1.884 | 1.71-2.08 | < 0.001 | 1.936 | 1.67-2.25 | < 0.001 |
| > 60 | 1.554 | 1.44-1.68 | < 0.001 | 1.598 | 1.42-1.80 | < 0.001 |
| Differentiation |  |  |  |  |  |  |
| G1 | 1.268 | 0.94-1.72 | 0.125 | 1.549 | 1.03-2.33 | 0.035 |
| G2 | 1.62 | 1.50-1.75 | < 0.001 | 1.678 | 1.49-1.89 | < 0.001 |
| G3 | 1.967 | 1.73-2.23 | < 0.001 | 1.859 | 1.53-2.26 | < 0.001 |
| Location |  |  |  |  |  |  |
| Rectum | 1.647 | 1.53-1.77 | < 0.001 | 1.613 | 1.44-1.81 | < 0.001 |
| Rectosigmoid junction | 1.73 | 1.55-1.93 | < 0.001 | 2.018 | 1.70-2.39 | < 0.001 |
| CEA |  |  |  |  |  |  |
| High | 1.735 | 1.56-1.93 | < 0.001 | 1.761 | 1.50-2.07 | < 0.001 |
| Normal | 1.701 | 1.52-1.91 | 0 | 1.7 | 1.42-2.02 | 0 |
| Preoperative RT |  |  |  |  |  |  |
| Yes | 1.522 | 1.38-1.68 | < 0.001 | 1.42 | 1.22-1.66 | < 0.001 |
| No | 1.782 | 1.65-1.93 | < 0.001 | 1.92 | 1.70-2.16 | < 0.001 |
| Chemotherapy |  |  |  |  |  |  |
| Yes | 1.68 | 1.56-1.81 | < 0.001 | 1.662 | 1.49-1.86 | < 0.001 |
| No | 1.62 | 1.44-1.82 | < 0.001 | 1.774 | 1.49-2.11 | < 0.001 |
| ELNs |  |  |  |  |  |  |
| 0-10 | 1.415 | 1.29-1.56 | < 0.001 | 1.439 | 1.24-1.67 | < 0.001 |
| 11-20 | 1.79 | 1.61-2.00 | < 0.001 | 2.10 | 1.78-2.48 | < 0.001 |
| > 20 | 3.30 | 2.64-4.14 | < 0.001 | 2.61 | 1.83-3.70 | < 0.001 |

1Log odds of negative lymph nodes/tumor size was treated as categorical variable, and the high group as reference; 2*P* values were adjusted for sex, age, race, tumor-node-metastasis stage, location, preoperative radiotherapy, carcinoembryonic antigen, differentiation and chemotherapy as covariates. CI: Confidence interval; CEA: Carcinoembryonic antigen; HR: Hazard ratio; RT: Radiotherapy; TNM: Tumor-node-metastasis; ELNs: Examined lymph nodes.