



Development and validation of a survival prediction model for 113,239 patients with colon cancer: a retrospective cohort study

Ying Li, Xiaorong Lai, Dongyang Yang, Dong Ma

Department II of Medical Oncology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China

Contributions: (I) Conception and design: Y Li, D Ma; (II) Administrative support: None; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: Y Li, D Yang; (V) Data analysis and interpretation: Y Li, X Lai; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Dong Ma. Department II of Medical Oncology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, 106 Zhongshan 2 Road, Guangzhou 510000, Guangdong Province, China. Email: Dongma_2003@outlook.com.

Background: Colon cancer (CC) is the third most commonly diagnosed malignant tumor and remains the second leading cause of cancer-related deaths worldwide. However, the risk assessment of poor prognosis of CC is limited in previous studies. This study aimed to develop a predictive nomogram for the survival of CC patients.

Methods: In this retrospective cohort study, 113,239 CC patients from the Surveillance, Epidemiology, and End Results (SEER) database were randomly divided into training (n=56,619) and testing (n=56,620) sets with a ratio of 1:1. Demographic, clinical data and survival status of patients were extracted. The outcomes were 3- and 5-year survival of CC. Univariate and multivariate Cox regression analyses were used to screen the predictors to develop the predictive nomogram. Internal validation and stratified analyses were further assessed the nomogram. The C-index and area under the curve (AUC) were calculated to estimate the model's predictive capacity, and calibration curves were adopted to estimate the model fit.

Results: Totally 38,522 (34.02%) patients died during the 5-year follow-up. The nomogram incorporated variables associated with the prognosis of CC patients, including age, gender, marital status, insurance status, tumor grade, stage (T/N/M), surgery, and number of nodes examined, with a C-index of 0.775 in the training set and 0.774 in the testing set. The AUCs of the nomogram for the 3- and 5-year survival prediction in the training set were 0.817 and 0.808, with the sensitivity of 0.688 and 0.716, and the specificity of 0.785 and 0.740, respectively. Similar results were found in the testing set. The C-index of the predictive nomogram for male, female, White, Black, and other races was 0.769, 0.779, 0.773, 0.770, and 0.770, respectively. The calibration curves for the nomogram in the above five cohorts showed a good agreement between actual and predicted values.

Conclusions: The nomogram may exhibit a certain predictive performance based on the SEER database, which may provide individual survival predictions for CC patients.

Keywords: Colon cancer (CC); survival; predictors; nomogram; Surveillance, Epidemiology, and End Results (SEER) database

Submitted Aug 22, 2022. Accepted for publication Oct 12, 2022.

doi: 10.21037/jgo-22-878

View this article at: <https://dx.doi.org/10.21037/jgo-22-878>

Introduction

Colon cancer (CC) is the third most commonly diagnosed malignant tumor and remains the second leading cause of cancer-related deaths worldwide. There were approximately

104,270 new cases of CC and 52,980 CC-related deaths in the United States in 2021 (1). Despite significant advances in the treatment of CC, its incidence continues to increase, and the 5-year survival rate remains low (2). Therefore, it

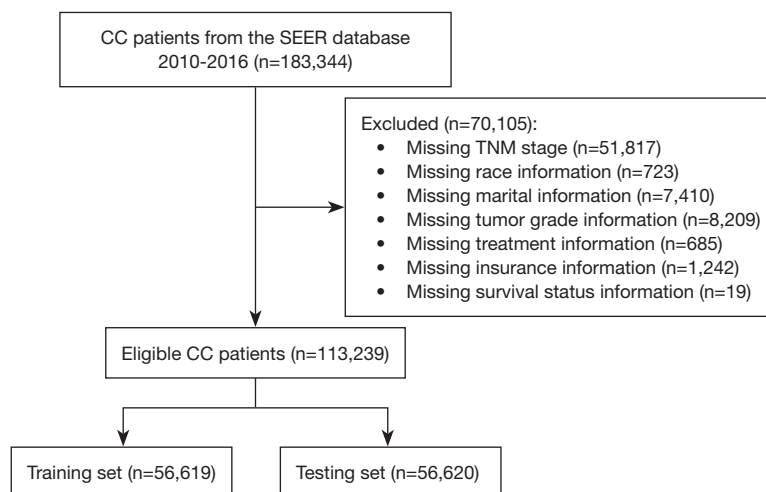


Figure 1 Flow chart of the selection process for CC patients. CC, colon cancer; SEER, Surveillance, Epidemiology, and End Results.

is essential for clinicians to identify the influencing factors associated with poor prognosis to improve the quality of life for CC patients.

Some clinicopathological characteristics have been associated with CC prognosis, such as age, race, and tumor site (3-5). Patients with proximal CC (right-sided CC) were shown to have a worse prognosis than those with distal CC (left-sided CC) (6). Several prediction models have been developed to predict the survival of CC patients. For example, a predictive model with 516 patients was constructed to assess the prognosis of CC patients (7). Nomograms can visualize complex regression equations to make the predictive results more intuitive and convenient for clinicians to use by integrating multiple clinical factors. Nomograms have been gradually applied in clinical research to predict the prognosis of various cancers, such as hepatic carcinoma and perihilar cholangiocarcinoma (8,9). Several prognostic nomograms have been proposed to assess the survival of CC patients, integrating stage and metastatic status using relatively small sample sizes (10,11). Zheng *et al.* developed a nomogram to predict the cancer-specific survival in 13,984 elderly patients with stages I–III CC, lacking relevant treatment information (10). Yu *et al.* conducted a prognostic nomogram to predict overall survival and cancer-specific survival among 11,220 old early-onset CC patients of age <50 years, similarly lacking the surgery and radiotherapy data (11). Furthermore, studies based on larger sample sizes with more stratified analyses are required for a further assessment of CC prognosis.

This study aimed to establish a nomogram to predict the

3- and 5-year survival of CC patients and to further assess the predictive performance of the nomogram using internal validation and stratified analyses (gender and race) in 113,239 participants from the Surveillance, Epidemiology, and End Results (SEER) database. We present the following article in accordance with the TRIPOD reporting checklist (available at <https://jgo.amegroups.com/article/view/10.21037/jgo-22-878/rc>).

Methods

Data sources

The data from CC cases in this retrospective cohort study were obtained from the SEER 18 Regs Custom Data (with additional treatment fields) of the National Cancer Institute (<http://seer.cancer.gov/>), which were collected from 2010 to 2016. The diagnosis of CC was confirmed using the International Classification of Diseases-Oncology 3 (ICD-O-3) 2008 site codes C180–C189 and C260. The SEER registries included data on patient demographics, primary tumor site, tumor morphology, stage at diagnosis, first course of treatment, and vital status of patients after follow-up. CC patients aged ≥ 18 years were included. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The detailed procedure for patient selection is presented in *Figure 1*.

Potential predictors

Demographic and clinical data of CC patients were

extracted from the SEER database, including age at diagnosis, gender, race (White, Black, and others), marital status (married, divorced, separated, single, unmarried or domestic partner, and widowed), insurance status (insured, insurance status unknown, any Medicaid, and uninsured), tumor grade, American Joint Committee on Cancer (AJCC) stage (T/N/M) (7th edition), number of nodes examined, treatments (surgery or radiotherapy).

Outcomes and follow-up

The 3- and 5-year survival of CC patients were defined as the outcomes. During the follow-up duration, the survival status of all patients was recorded. And the follow-up was terminated when the patient died.

Development and validation of the nomogram

All CC patients were randomly divided into the training and testing sets with a 1:1 ratio, which has ensured that the variables of the two sets can be balanced. The predictor screening of the survival of CC patients was conducted using the training set, and then the prediction model was performed based on the predictors. The internal validation of the model was conducted using the testing set. A visual nomogram was drawn with the predictors. The performance of the prediction model was assessed by C-index, area under the curve (AUC), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). Receiver operator characteristic (ROC) curve was used to evaluate the discernibility ability of the model. Calibration curves were adopted to estimate the model fit.

Statistical analysis

All statistical analyses were performed by SAS 9.4 (SAS Institute, Cary, NC, USA) and R software 4.0.3 (R Foundation for Statistical Computing, Vienna, Austria). Count data are described by the number of cases/constituent ratio [n (%)], and the χ^2 test or Fisher's exact test was adopted for intergroup comparisons. Univariate and multivariate Cox regression analyses were conducted to identify the prognostic factors. The included patients were randomly divided into a training set (n=56,619) and a testing set (n=56,620). Stratified analyses were conducted in terms of gender and race. Hazard ratio (HRs) and 95% confidence interval (CIs) were

calculated. Statistical significance levels were all two-sided. A P value <0.05 was considered statistically significant.

Results

Characteristics of the study population in the training cohort

A total of 183,344 CC patients were initially obtained from the SEER database. After excluding 51,817 patients with incomplete TNM stage information, 8,133 patients with missing race and marital information, 8,209 patients with unknown tumor grades, 685 patients with missing the treatment data, 1,242 patients with unknown insurance information, and 19 patients with unknown survival status, 113,239 participants were eventually included in the study (Figure 1). The study population's characteristics in the training cohort are shown in Table 1. In total, 56,619 CC cases were finally enrolled in the training set, with an average age of 67.85 years. Of these, 27,902 (49.28%) were male and 28,717 (50.72%) were female. The study population was ethnically diverse and comprised White (45,126/79.70%), Black (6,761/11.94%), and other races (4,732/8.36%). There were 30,961 (54.68%) married patients, 5,625 (9.93%) divorced patients, 601 (1.06%) separated patients, 9,023 (15.94%) single patients, 141 (0.25%) unmarried patients or those with a domestic partner, and 10,268 (18.14%) widowed patients. For insurance status, there were 38,526 (68.04%) insured patients, 9,459 (16.71%) patients without insurance status unknown, 6,987 (12.34%) patients receiving any Medicaid, and 1,647 (2.91%) patients without insurance. The number of patients with grade I, II, III, and IV tumors was 5,754 (10.16%), 38,943 (68.78%), 9,732 (17.19%), and 2,190 (3.87%), respectively. There were 9,357 (16.91%) patients with T1 tumors, 7,801 patients (13.78%) with T2 tumors, 28,669 patients (50.63%) with T3 tumors, and 10,792 patients (19.06%) with T4 tumors. There were 33,494 (59.16%) patients with N0 tumors, 14,216 (25.11%) with N1 tumors, and 8,909 (15.45%) with N2 tumors. Additionally, 47,871 (84.55%) patients had M0 tumors, and 8,748 (15.45%) had M1 tumors. Regarding treatments, 54,638 (96.50%) had surgery, and 1,981 (3.50%) had none; 990 (1.75%) underwent radiotherapy, and 55,629 (98.25%) had none. The median number of nodes examined was 16, and the median survival time was 30 months.

Table 1 Baseline characteristics of patients in the training cohort

Variables	CC patients (n=56,619)
Age, years, mean \pm SD	67.85 \pm 14.06
Gender, n (%)	
Male	27,902 (49.28)
Female	28,717 (50.72)
Race, n (%)	
White	45,126 (79.70)
Black	6,761 (11.94)
Others	4,732 (8.36)
Marital status, n (%)	
Married	30,961 (54.68)
Divorced	5,625 (9.93)
Separated	601 (1.06)
Single	9,023 (15.94)
Unmarried or domestic partner	141 (0.25)
Widowed	10,268 (18.14)
Insurance status, n (%)	
Insured	38,526 (68.04)
Insurance status unknown	9,459 (16.71)
Any Medicaid	6,987 (12.34)
Uninsured	1,647 (2.91)
Tumor grade, n (%)	
I	5,754 (10.16)
II	38,943 (68.78)
III	9,732 (17.19)
IV	2,190 (3.87)
T stage, n (%)	
T1	9,357 (16.91)
T2	7,801 (13.78)
T3	28,669 (50.63)
T4	10,792 (19.06)
N stage, n (%)	
N0	33,494 (59.16)
N1	14,216 (25.11)
N2	8,909 (15.45)

Table 1 (Continued)**Table 1** (Continued)

Variables	CC patients (n=56,619)
M stage, n (%)	
M0	47,871 (84.55)
M1	8,748 (15.45)
Surgery, n (%)	54,638 (96.50)
Radiotherapy, n (%)	990 (1.75)
Nodes examined, M (Q ₁ , Q ₃)	16.00 (12.00, 23.00)
Survival time, months, M (Q ₁ , Q ₃)	30.00 (15.00, 52.00)

CC, colon cancer; SD, standard deviation.

Predictor screening for the survival of CC patients in the training cohort

Predictor screening for the survival of CC patients in the training cohort are exhibited in *Table 2*. Factors with significant differences in the univariate Cox analysis were further analyzed using multivariate Cox regression and included age, gender, race, marital status, insurance status, tumor grade, stage (T/N/M), surgery, and number of nodes examined. The outcomes showed that increased age was associated with poorer survival (HR =1.043, 95% CI: 1.041–1.044, P<0.05). An increased number of nodes examined was associated with a decreased risk of poor survival (HR =0.982, 95% CI: 0.981–0.984, P<0.05). Female patients had better survival than males (HR =0.834, 95% CI: 0.810–0.859, P<0.05). In terms of race, Black patients (HR =1.090, 95% CI: 1.044–1.138, P<0.05) had a worse survival than White patients, whereas other races (HR =0.837, 95% CI: 0.792–0.884, P<0.05) had a reduced risk of poor survival compared with White patients. For marital status, divorced (HR =1.219, 95% CI: 1.162–1.280, P<0.05), single (HR =1.332, 95% CI: 1.277–1.389, P<0.05), and widowed (HR =1.261, 95% CI: 1.213–1.311, P<0.05) patients had a shorter survival time than married patients. Additionally, patients with no specific insured status (HR =1.091, 95% CI: 1.052–1.132, P<0.05), any Medicaid (HR =1.377, 95% CI: 1.319–1.437, P<0.05), or who were uninsured (HR =1.451, 95% CI: 1.331–1.580, P<0.05) had a worse survival than insured patients. Patients with grade II (HR =1.225, 95% CI: 1.156–1.298, P<0.05), grade III (HR =1.561, 95% CI: 1.465–1.664, P<0.05), and grade IV tumors (HR =1.795, 95% CI: 1.654–1.947, P<0.05) had a higher risk of poor

Table 2 Predictor screening for the survival of CC patients

Variables	Univariable Cox regression model		Multivariate Cox regression model	
	HR (95% CI)	P	HR (95% CI)	P
Age	1.032 (1.031, 1.033)	<0.001	1.043 (1.041, 1.044)	<0.001
Gender				
Male	Ref		Ref	
Female	0.965 (0.938, 0.992)	0.011	0.834 (0.810, 0.859)	<0.001
Race				
White	Ref		Ref	
Black	1.059 (1.015, 1.105)	0.008	1.090 (1.044, 1.138)	<0.001
Others	0.822 (0.779, 0.869)	<0.001	0.837 (0.792, 0.884)	<0.001
Marital status				
Married	Ref		Ref	
Divorced	1.251 (1.192, 1.312)	<0.001	1.219 (1.162, 1.280)	<0.001
Separated	1.024 (0.887, 1.183)	0.744	1.065 (0.921, 1.231)	0.394
Single	1.216 (1.168, 1.266)	<0.001	1.332 (1.277, 1.389)	<0.001
Unmarried or domestic partner	0.953 (0.680, 1.334)	0.779	1.273 (0.909, 1.783)	0.160
Widowed	1.800 (1.740, 1.864)	<0.001	1.261 (1.213, 1.311)	<0.001
Insurance status				
Insured	Ref		Ref	
Insurance status unknown	1.248 (1.203, 1.294)	<0.001	1.091 (1.052, 1.132)	<0.001
Any Medicaid	1.366 (1.311, 1.423)	<0.001	1.377 (1.319, 1.437)	<0.001
Uninsured	1.109 (1.020, 1.205)	0.015	1.451 (1.331, 1.580)	<0.001
Tumor grade				
I	Ref		Ref	
II	1.497 (1.414, 1.585)	<0.001	1.225 (1.156, 1.298)	<0.001
III	2.597 (2.442, 2.762)	<0.001	1.561 (1.465, 1.664)	<0.001
IV	2.966 (2.739, 3.212)	<0.001	1.795 (1.654, 1.947)	<0.001
T stage				
T1	Ref		Ref	
T2	0.845 (0.793, 0.901)	<0.001	0.954 (0.893, 1.019)	0.163
T3	1.480 (1.414, 1.550)	<0.001	1.269 (1.205, 1.336)	<0.001
T4	3.225 (3.072, 3.386)	<0.001	1.974 (1.868, 2.086)	<0.001
N stage				
N0	Ref		Ref	
N1	1.686 (1.630, 1.743)	<0.001	1.397 (1.348, 1.447)	<0.001
N2	2.972 (2.872, 3.076)	<0.001	2.151 (2.066, 2.239)	<0.001

Table 2 (Continued)

Table 2 (Continued)

Variables	Univariable Cox regression model		Multivariate Cox regression model	
	HR (95% CI)	P	HR (95% CI)	P
M stage				
M0	Ref		Ref	
M1	4.445 (4.313, 4.581)	<0.001	3.106 (2.998, 3.219)	<0.001
Surgery				
Yes	Ref		Ref	
No	5.286 (5.019, 5.567)	<0.001	2.930 (2.753, 3.119)	<0.001
Radiotherapy				
Yes	Ref			
No	0.691 (0.631, 0.757)	<0.001		
Nodes examined	0.976 (0.975, 0.978)	<0.001	0.982 (0.981, 0.984)	<0.001

CC, colon cancer; HR, hazard ratio; CI, confidence interval; ref: reference.

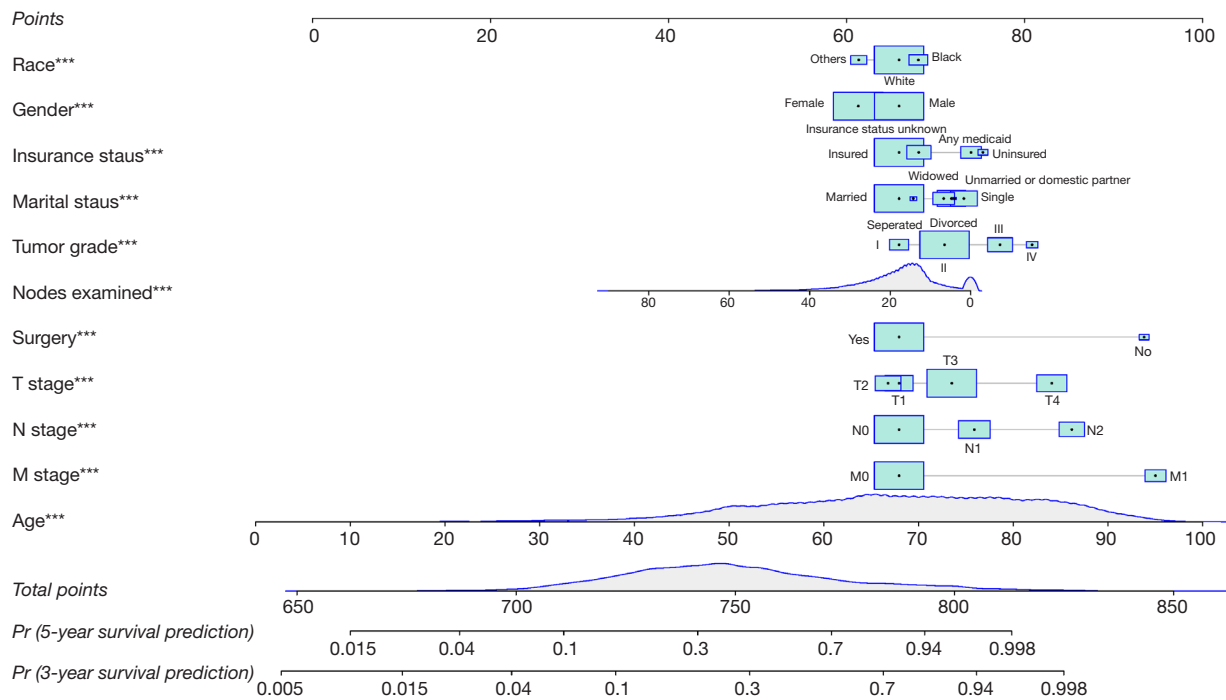


Figure 2 Nomogram for predicting the 3- and 5-year survival of CC patients. ***, P<0.001. CC, colon cancer.

survival than patients with grade I tumors. As for T stage, patients with T3 (HR =1.269, 95% CI: 1.205–1.336, P<0.05) and T4 tumors (HR =1.974, 95% CI: 1.868–2.086, P<0.05) had a worse survival than patients with T1 tumors. Individuals with N1 (HR =1.397, 95% CI: 1.348–1.447, P<0.05) and N2 tumors (HR =2.151, 95% CI: 2.066–2.239, P<0.05) had a

shorter survival time than those with N0 tumors. The survival rate of patients with M1 tumors was lower than patients with M0 tumors (HR =3.106, 95% CI: 2.998–3.219, P<0.05). Patients without surgery had worse survival than patients with surgery (HR =2.930, 95% CI: 2.753–3.119, P<0.05). Based on these predictors, a nomogram was established to predict the 3-

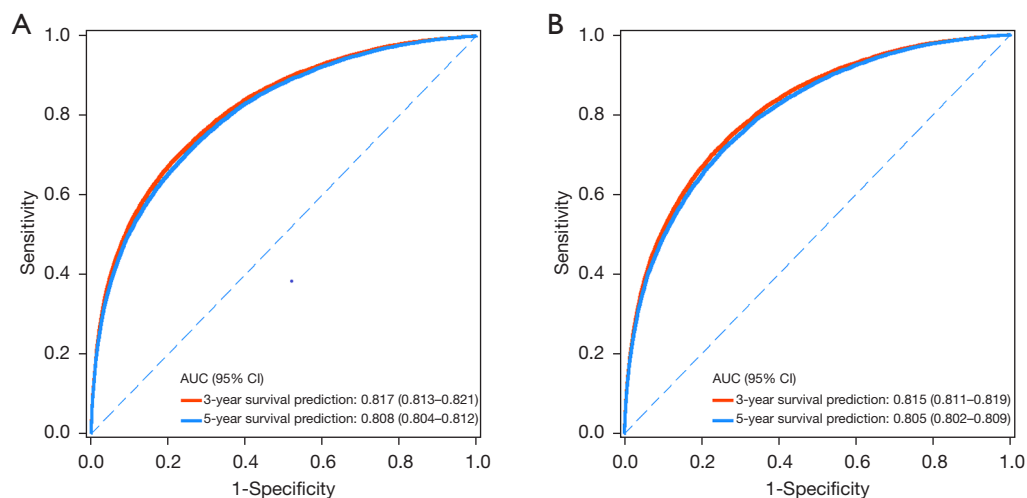


Figure 3 ROC curves of the model for predicting the 3- and 5-year survival of CC patients. (A) Training cohort. (B) Testing cohort. ROC, receiver operator characteristic; AUC, area under the curve; CC, colon cancer.

Table 3 The predictive performance of the model for 3- and 5-year survival

Model	AUC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	NPV (95% CI)	PPV (95% CI)	Accuracy (95% CI)
Training set						
3-year survival prediction	0.817 (0.813–0.821)	0.688 (0.681–0.695)	0.785 (0.781–0.789)	0.861 (0.858–0.865)	0.565 (0.558–0.572)	0.757 (0.754–0.761)
5-year survival prediction	0.808 (0.804–0.812)	0.716 (0.710–0.723)	0.740 (0.736–0.745)	0.836 (0.832–0.840)	0.586 (0.579–0.592)	0.732 (0.729–0.736)
Testing set						
3-year survival prediction	0.815 (0.811–0.819)	0.688 (0.681–0.695)	0.781 (0.777–0.785)	0.860 (0.856–0.863)	0.563 (0.556–0.570)	0.754 (0.751–0.758)
5-year survival prediction	0.805 (0.802–0.809)	0.716 (0.710–0.723)	0.740 (0.735–0.744)	0.834 (0.830–0.838)	0.588 (0.582–0.595)	0.732 (0.728–0.735)

AUC, area under the curve; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value.

and 5-year survival for CC patients (Figure 2).

A Cox regression model was established as follows: $Y = 0.042 \text{ age} - 0.181 \text{ female} + 0.086 \text{ Black} - 0.178 \text{ other} + 0.198 \text{ married} + 0.286 \text{ single} + 0.232 \text{ widowed} + 0.087 \text{ insured} + 0.320 \text{ any Medicaid} + 0.372 \text{ uninsured} + 0.203 \text{ grade II} + 0.445 \text{ grade III} + 0.585 \text{ grade IV} + 0.026 \text{ T3 stage} + 0.680 \text{ T4 stage} + 0.334 \text{ N2 stage} + 0.766 \text{ N3 stage} + 1.133 \text{ M stage} + 1.075 \text{ surgery} - 0.018 \text{ examined nodes}$.

Development and validation of the nomogram

The ROC curves of the prediction model are shown in Figure 3. The AUC of the model for predicting the 3- and 5-year survival of CC patients are displayed in Table 3.

In the training set, the AUC of the model of the 3-year survival prediction was 0.817 (95% CI: 0.813–0.821), with the sensitivity of 0.688 (95% CI: 0.681–0.695) and the specificity of 0.785 (95% CI: 0.781–0.789). And the AUC of the model of the 5-year survival prediction was 0.808 (95% CI: 0.804–0.812), with the sensitivity of 0.716 (95% CI: 0.710–0.723) and the specificity of 0.740 (95% CI: 0.736–0.745). In the testing set, the AUC of the model of the 3- and 5-year survival prediction were 0.815 (95% CI: 0.811–0.819) and 0.805 (95% CI: 0.802–0.809), respectively.

The predictive performance of the nomogram for predicting the 3- and 5-year survival are presented in Table 4. The C-index for the predictive nomogram was 0.775 (95% CI: 0.771–0.779) in the training set and was confirmed to

Table 4 The predictive performance of the nomogram

Groups	C-index	S.E.	95% CI
Set			
Training	0.775	0.002	0.771, 0.779
Testing	0.774	0.002	0.770, 0.778
Gender			
Male	0.769	0.002	0.765, 0.773
Female	0.779	0.002	0.775, 0.783
Race			
White	0.773	0.002	0.769, 0.777
Black	0.770	0.005	0.760, 0.780
Others	0.770	0.005	0.760, 0.780

CI, confidence interval; S.E., standard error.

be 0.774 (95% CI: 0.770–0.778) in the testing set. The C-index for the nomogram in male patients and female patients was 0.769 (95% CI: 0.765–0.773) and 0.779 (95% CI: 0.775–0.783), respectively. In addition, the C-index for the nomogram in White, Black, and other races were 0.773 (95% CI: 0.769–0.777), 0.770 (95% CI: 0.760–0.780), and 0.770 (95% CI: 0.760–0.780), respectively.

The calibration curves for survival predicted by the nomogram in different cohorts are illustrated in *Figures 3–5*, including the training (*Figure 4A*) and testing (*Figure 4B*) sets; male patients (*Figure 5A*); female patients (*Figure 5B*); and White (*Figure 6A*), Black (*Figure 6B*), and other races (*Figure 6C*). The results showed a good agreement between the actual and predicted values.

Sample

A patient's information, randomly selected from the training set, is used as an example: 85 years old, White race, male, married, insured, tumor grade II, T2 stage, N0 stage, M0 stage, history of surgery, no radiation therapy, and 16 nodes examined. The patient's total score calculated by the nomogram was 747 points; the 3-year risk of death was 0.241; the 5-year risk of death was 0.357 (*Figure 7*). The patient's actual survival status is "survival" indicating that the nomogram-predicted results were correct in this case.

Discussion

CC is one of the most common types of malignant tumors

worldwide, posing a severe threat to human life and health. According to statistics, the survival of CC patients is estimated to be 50–70%, with a 5-year net survival lower than 50% in some countries (12). In addition, the prognosis of CC is influenced by many complex factors, increasing the difficulty of evaluating prognosis and making therapeutic decisions. Therefore, to improve the prognosis of CC patients, it is essential for physicians to identify patients with a poor prognosis. This study developed and validated a nomogram to predict the 3- and 5-year survival of CC patients based on 113,239 participants. The nomogram incorporated variables associated with CC prognosis, including age, gender, race, marital status, insurance status, tumor grade, stage (T/N/M), surgery, and number of nodes examined. The C-index of the nomogram was 0.775 (95% CI: 0.771–0.779) in the training set and confirmed as 0.774 (95% CI: 0.770–0.778) in the testing set. In addition, the C-index for the predictive nomogram in White, Black and other races was 0.773 (95% CI: 0.769–0.777), 0.770 (95% CI: 0.760–0.780), and 0.770 (95% CI: 0.760–0.780), respectively. The calibration curves in the training and testing cohorts suggested the nomogram may have a good predictive ability. These findings indicated that the developed nomogram had a certain ability to predict the survival of CC patients.

Our results showed that marital status affected CC prognosis, which could be attributed to the following plausible explanations: (I) married people are more likely to develop healthier lifestyles with the supervision and help of their spouses (13); (II) patients with bad marriages or single patients are more vulnerable to negative emotions, such as anxiety and hopelessness, adversely affecting individual coping strategies when facing cancer (14); (III) spouses may encourage patients to receive treatments positively and provide practical assistance and care (15). Age was used as a prognostic indicator for CC patients. Our findings indicated that the prognosis for older patients with CC was worse than that for younger patients. For elderly CC patients, a full evaluation is needed when selecting surgical resections that may cause significant trauma. Strategies for enhancing geriatric care might help to reduce the poor outcome of elderly CC patients. Of note, we observed that CC patients from the SEER database were mostly over 60 years old. Since the onset age of CC tends to be younger (16), the relationship between age and prognosis should be interpreted with caution. Concerning gender, we found that male patients with CC had a poorer prognosis than female patients, probably partly due to differences in hormone

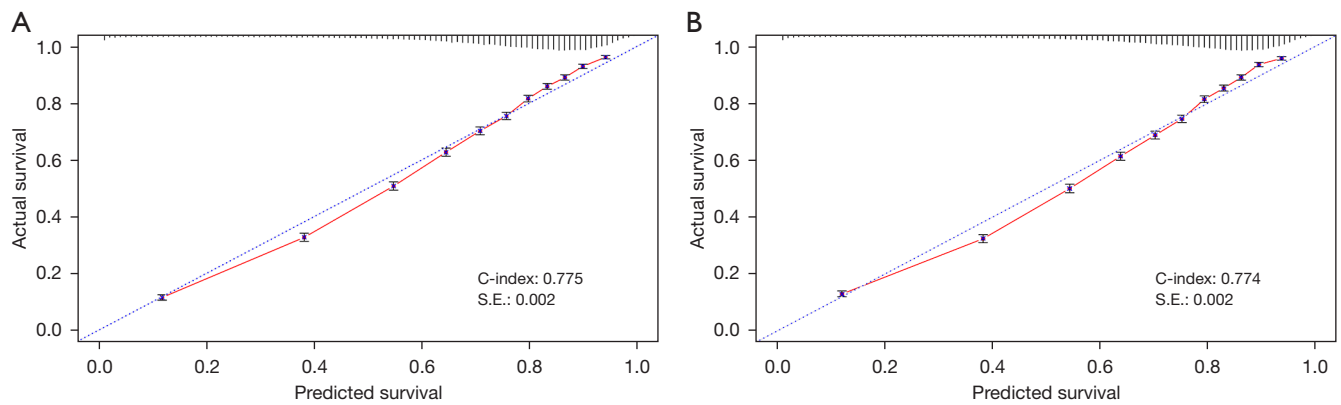


Figure 4 Calibration plots of the nomogram prediction in CC patients. (A) Training cohort. (B) Testing cohort. CC, colon cancer.

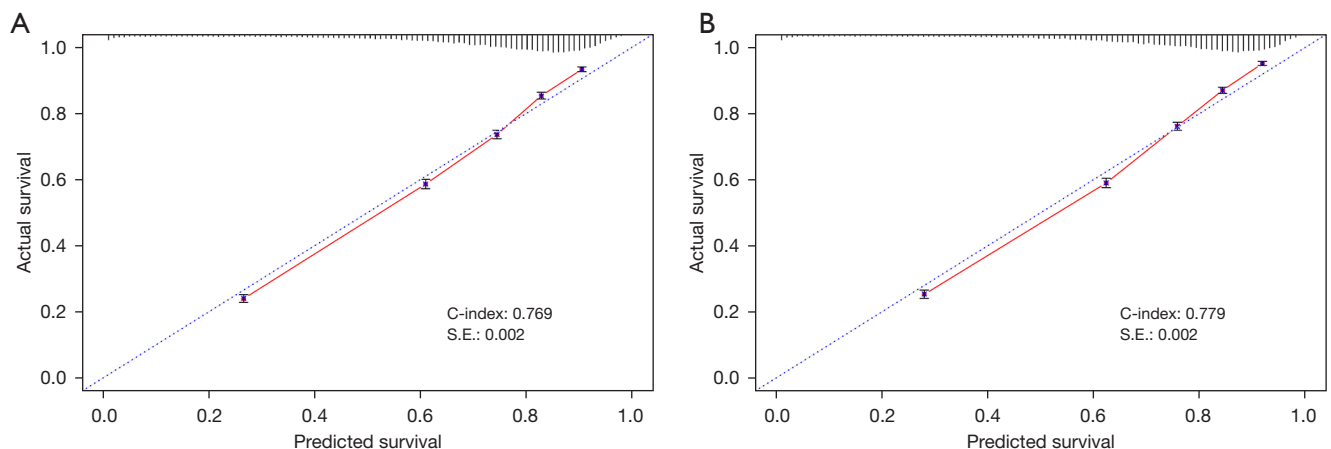


Figure 5 Calibration plots of the nomogram prediction in CC patients. (A) Males. (B) Females. CC, colon cancer.

levels. A recent study indicated that higher androgen levels were associated with the formation of CC tumors because they promote faster intestinal stem cell division and induce a decreased production of mature epithelial cells (17).

Previous studies have found that CC patients with Medicaid only or no insurance had worse survival, consistent with our results (18,19). Patients without insurance were less inclined to receive cancer screening, and most had already progressed to severe stages when diagnosed (20,21). Additionally, they may accept an incomplete assessment of symptoms and refuse chemotherapy for their advanced disease (22). The present study found that the prognosis of CC cases differed in diverse races, with a better prognosis seen in White patients than Black patients. Sineshaw *et al.* revealed that Black Americans with CC had higher mortality than Caucasians due to the differences in insurance (23).

Full implementation of healthcare security policies might be a vital measure to improve CC patients' prognosis.

The TNM staging system has been widely used in the pathological staging of CC. T stage refers to the depth of the tumor invasion, affecting the prognosis of CC patients. Lymphatic vessels of the colonic wall arise from the submucosal layer, and if the tumor infiltration reaches the submucosal layer, lymph node metastasis may occur (24-27). The deeper the invasion, the higher the probability of lymph node metastasis and the poorer the prognosis. In most cases, tumor cell infiltration of surrounding tissues by breaking through the serosal surface was observed in the T4 stage. Additionally, intestinal obstruction and intestinal perforation generally occur in patients with a T4-stage tumor. Burdy *et al.* suggested that the T4 stage could independently predict the prognosis of CC patients

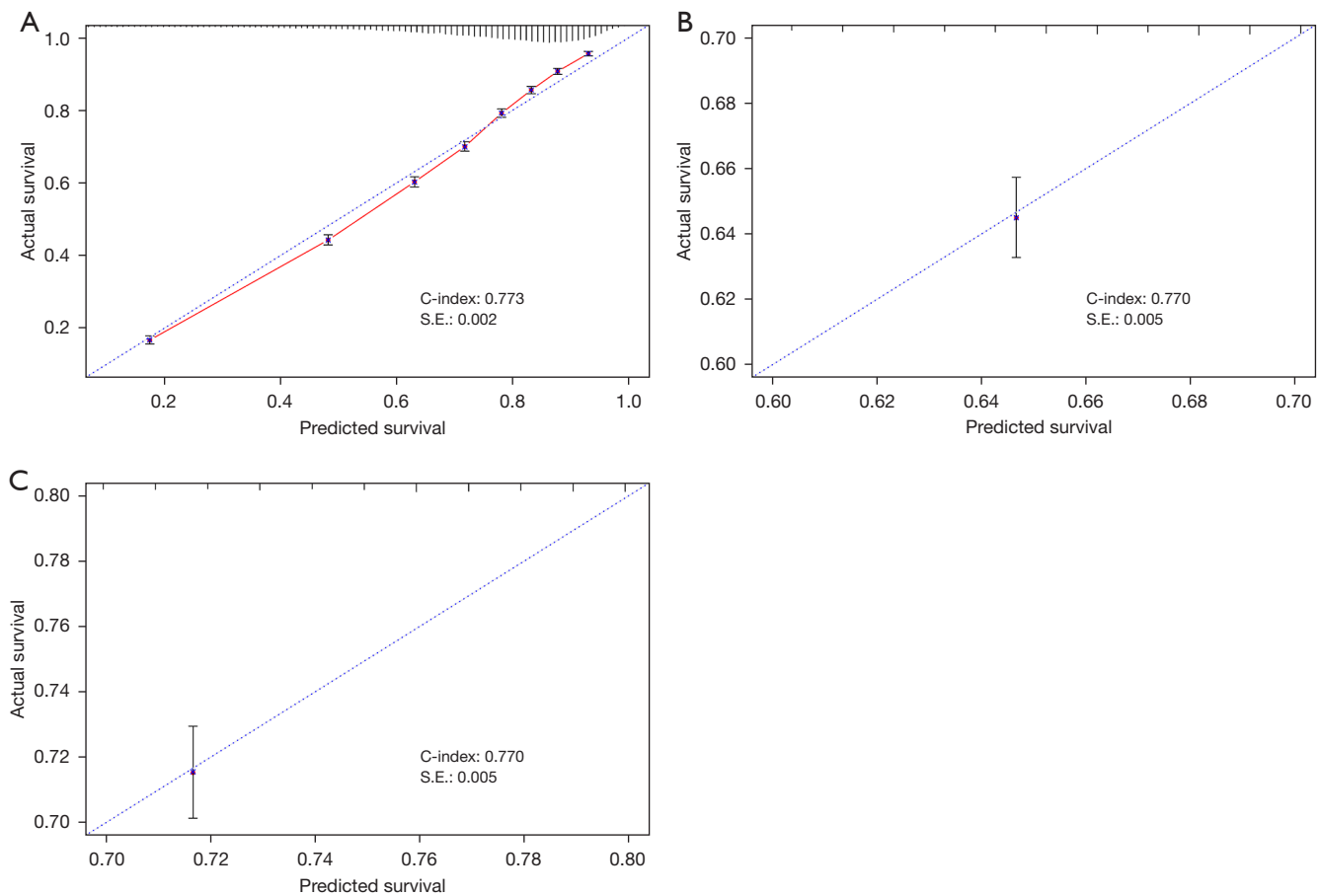


Figure 6 Calibration plots of the nomogram prediction in CC patients. (A) White race (B) Black race. (C) Other races. CC, colon cancer.

with grade II tumors who needed to receive postoperative adjuvant chemotherapy (28). Prior research has demonstrated that lymph node metastasis was also an independent predictor of prognosis for CC patients. The greater the number of lymph node metastases, the worse the prognosis, and patients with lymph node metastases had a poorer prognosis than those without lymph node metastases (29). Adjuvant chemotherapy was adopted for these patients (30,31).

The number of nodes examined has become one of the most critical indicators affecting the surgery and medical treatment of CC patients. An adequate number of nodes examined might contribute to a more accurate diagnosis of CC to guide adjuvant therapy. Consistent with our results, Ramser *et al.* also found that the number of nodes examined was an independent predictor of CC prognosis. Detecting more than or equal to 12 lymph nodes could improve the prognosis of patients with CC, indicating that too few nodes examined was associated with a poor prognosis, especially

in patients with grade II tumors (32). Clinically, fewer nodes examined could result in an inaccurate pathological assessment, which might influence subsequent adjuvant therapy, and further affect prognosis.

Although several nomograms have been used for CC prognosis prediction (33,34), the small sample size of these studies may be insufficient for clinicians to apply practically. In this study, a multi-factor prediction system was established based on 113,239 participants, which may help clinicians identify CC patients with a poorer 3- and 5-year survival. In clinical practice, physicians should not only consider the severity of the patient's illness from the clinical perspective but also consider the patient's insurance status from the perspective of health economics to accurately assess prognosis. Herein, the multi-dimensional and personalized quantitative model proposed by this study will help physicians estimate the survival of CC patients and identify the populations who need close follow-up.

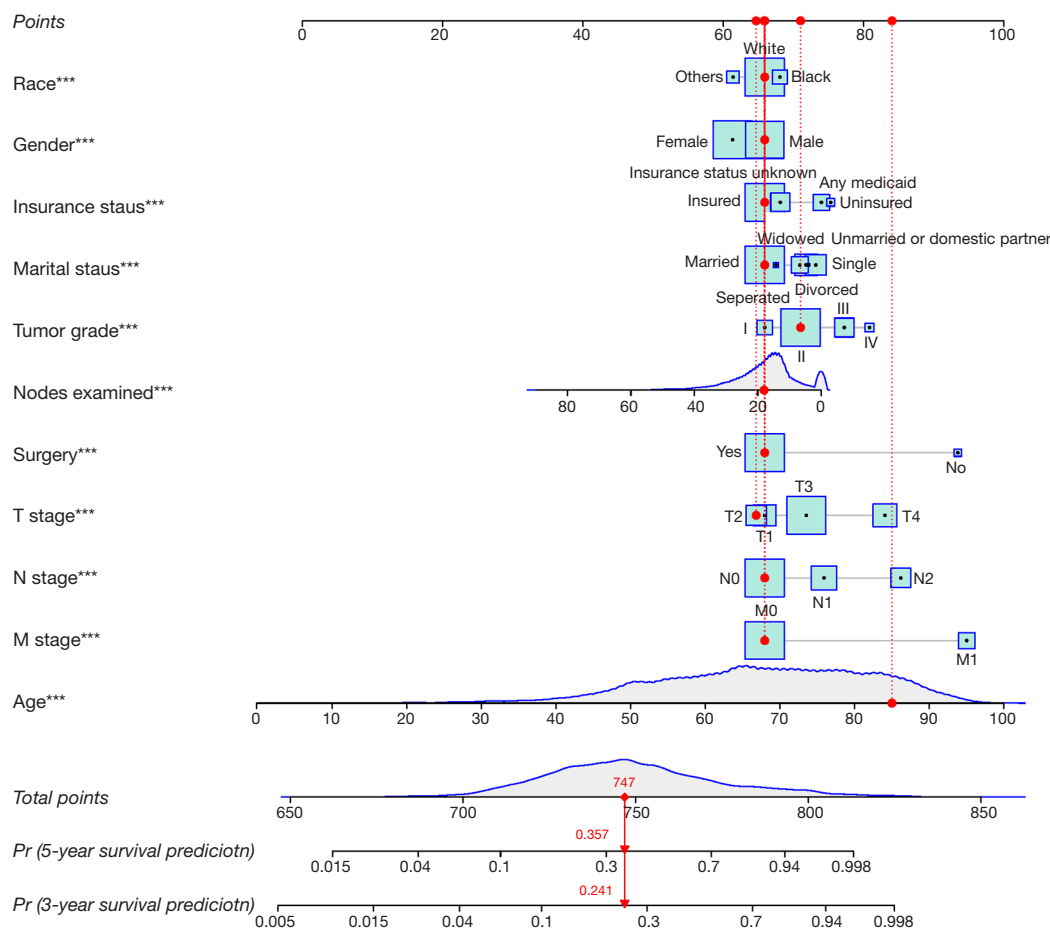


Figure 7 Results of the nomogram in predicting the survival of a patient randomly selected from the training set. The patient's information: 85 years old, White race, male, married, insured, tumor grade II, T2 stage, N0 stage, M0 stage, history of surgery, no radiation, 16 nodes examined. ***, $P < 0.001$.

Several limitations in this study cannot be ignored. Firstly, the main limitation of our nomogram was the absence of external validation. Secondly, the lack of biological markers and laboratory indicators might affect the prediction performance of the nomogram. Thirdly, the effect of different surgical procedures on prognosis could not be analyzed, which should be assessed in future studies. Additionally, information on adjuvant therapy for CC patients with lymph node metastasis was not available from the SEER database, which may have influenced the predictive accuracy of the nomogram. Further models could be built based on a combination of clinical features and biological markers, with external validation to achieve a more accurate survival predictive performance in CC patients.

Conclusions

We developed a nomogram that performed well in predicting survival for 113,239 participants from the SEER database, which may prove helpful in providing individual survival predictions for CC patients.

Acknowledgments

Funding: None.

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at <https://jgo>.

amegroups.com/article/view/10.21037/jgo-22-878/rc

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://jgo.amegroups.com/article/view/10.21037/jgo-22-878/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Available online: <https://www.cancer.org/cancer/colon-rectal-cancer/about/key-statistics.html>
2. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 2021;71:209-49.
3. Huang Y, Ji L, Zhu J, et al. Lymph node status and its impact on the prognosis of left-sided and right-sided colon cancer: A SEER population-based study. *Cancer Med* 2021;10:8708-19.
4. Wang Y, Liu J, Ren F, et al. Identification and Validation of a Four-Long Non-coding RNA Signature Associated With Immune Infiltration and Prognosis in Colon Cancer. *Front Genet* 2021;12:671128.
5. Feng H, Lyu Z, Zheng J, et al. Association of tumor size with prognosis in colon cancer: A Surveillance, Epidemiology, and End Results (SEER) database analysis. *Surgery* 2021;169:1116-23.
6. Rumpold H, Hackl M, Petzer A, et al. Improvement in colorectal cancer outcomes over time is limited to patients with left-sided disease. *J Cancer Res Clin Oncol* 2022. [Epub ahead of print]. doi: 10.1007/s00432-021-03868-0.
7. Cai HJ, Zhuang ZC, Wu Y, et al. Development and validation of a ferroptosis-related lncRNAs prognosis signature in colon cancer. *Bosn J Basic Med Sci* 2021;21:569-76.
8. Zhang Y, Lei X, Xu L, et al. Preoperative and postoperative nomograms for predicting early recurrence of hepatocellular carcinoma without macrovascular invasion after curative resection. *BMC Surg* 2022;22:233.
9. Yu Z, Liu Q, Liao H, et al. Prognostic nomogram for predicting cancer-specific survival in patients with resected hilar cholangiocarcinoma: a large cohort study. *J Gastrointest Oncol* 2022;13:833-46.
10. Zheng P, Lai C, Yang W, et al. Nomogram predicting cancer-specific survival in elderly patients with stages I-III colon cancer. *Scand J Gastroenterol* 2020;55:202-8.
11. Yu C, Zhang Y. Development and validation of a prognostic nomogram for early-onset colon cancer. *Biosci Rep* 2019;39:BSR20181781.
12. Allemani C, Matsuda T, Di Carlo V, et al. Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *Lancet* 2018;391:1023-75.
13. Umberson D, Thomeer MB. Family Matters: Research on Family Ties and Health, 2010-2020. *J Marriage Fam* 2020;82:404-19.
14. Secinti E, Tometich DB, Johns SA, et al. The relationship between acceptance of cancer and distress: A meta-analytic review. *Clin Psychol Rev* 2019;71:27-38.
15. Buja A, Lago L, Lago S, et al. Marital status and stage of cancer at diagnosis: A systematic review. *Eur J Cancer Care (Engl)* 2018;27.
16. Friedenreich CM, Shaw E, Neilson HK, et al. Epidemiology and biology of physical activity and cancer recurrence. *J Mol Med (Berl)* 2017;95:1029-41.
17. Yu X, Li S, Xu Y, et al. Androgen Maintains Intestinal Homeostasis by Inhibiting BMP Signaling via Intestinal Stromal Cells. *Stem Cell Reports* 2020;15:912-25.
18. Hao S, Snyder RA, Irish W, et al. Association of race and health insurance in treatment disparities of colon cancer: A retrospective analysis utilizing a national population database in the United States. *PLoS Med* 2021;18:e1003842.
19. Zhou C, Zhang Y, Hu X, et al. The effect of marital and insurance status on the survival of elderly patients with stage M1b colon cancer: a SEER-based study. *BMC Cancer* 2021;21:891.

20. Bandi P, Minihan AK, Siegel RL, et al. Updated Review of Major Cancer Risk Factors and Screening Test Use in the United States in 2018 and 2019, with a Focus on Smoking Cessation. *Cancer Epidemiol Biomarkers Prev* 2021;30:1287-99.
 21. Unger JM, Blanke CD, LeBlanc M, et al. Association of Patient Demographic Characteristics and Insurance Status With Survival in Cancer Randomized Clinical Trials With Positive Findings. *JAMA Netw Open* 2020;3:e203842.
 22. Davis RE, Trickey AW, Abrahamse P, et al. Association of Cumulative Social Risk and Social Support With Receipt of Chemotherapy Among Patients With Advanced Colorectal Cancer. *JAMA Netw Open* 2021;4:e2113533.
 23. Sineshaw HM, Ng K, Flanders WD, et al. Factors That Contribute to Differences in Survival of Black vs White Patients With Colorectal Cancer. *Gastroenterology* 2018;154:906-915.e7.
 24. Hu S, Li S, Teng D, et al. Analysis of risk factors and prognosis of 253 lymph node metastasis in colorectal cancer patients. *BMC Surg* 2021;21:280.
 25. Xu Y, Chen Y, Long C, et al. Preoperative Predictors of Lymph Node Metastasis in Colon Cancer. *Front Oncol* 2021;11:667477.
 26. Shao YJ, Ni JJ, Wei SY, et al. IRF1-mediated immune cell infiltration is associated with metastasis in colon adenocarcinoma. *Medicine (Baltimore)* 2020;99:e22170.
 27. Deng J, Zhou S, Wang Z, et al. Comparison of Prognosis and Lymph Node Metastasis in T1-Stage Colonic and Rectal Carcinoma: A Retrospective Study. *Int J Gen Med* 2022;15:3651-62.
 28. Burdy G, Panis Y, Alves A, et al. Identifying patients with T3-T4 node-negative colon cancer at high risk of recurrence. *Dis Colon Rectum* 2001;44:1682-8.
 29. Lee SY, Lee J, Park HM, et al. Perineural invasion and number of retrieved lymph nodes are prognostic factors for T2N0 colon cancer. *Langenbecks Arch Surg* 2021;406:1979-85.
 30. Liu FQ, Cai SJ. Adjuvant and perioperative neoadjuvant therapy for colorectal cancer. *Zhonghua Wei Chang Wai Ke Za Zhi* 2019;22:315-20.
 31. Yeom SS, Lee SY, Kim CH, et al. The prognostic effect of adjuvant chemotherapy in the colon cancer patients with solitary lymph node metastasis. *Int J Colorectal Dis* 2019;34:1483-90.
 32. Ramser M, Lobbes LA, Warschkow R, et al. Evaluation of the prognostic relevance of the recommended minimum number of lymph nodes in colorectal cancer—a propensity score analysis. *Int J Colorectal Dis* 2021;36:779-89.
 33. Chen H, Luo J, Guo J. Development and validation of a five-immune gene prognostic risk model in colon cancer. *BMC Cancer* 2020;20:395.
 34. Tong D, Tian Y, Zhou T, et al. Improving prediction performance of colon cancer prognosis based on the integration of clinical and multi-omics data. *BMC Med Inform Decis Mak* 2020;20:22.
- (English Language Editor: D. Fitzgerald)

Cite this article as: Li Y, Lai X, Yang D, Ma D. Development and validation of a survival prediction model for 113,239 patients with colon cancer: a retrospective cohort study. *J Gastrointest Oncol* 2022;13(5):2393-2405. doi: 10.21037/jgo-22-878