

Development and external validation of a nomogram predicting overall survival after curative resection of colon cancer

Shuanhu Wang¹ , Yakui Liu¹, Yi Shi¹,
Jiajia Guan¹, Mulin Liu^{1,*}  and
Wenbin Wang^{2,*}

Abstract

Objective: To develop and externally validate a prognostic nomogram to predict overall survival (OS) in patients with resectable colon cancer.

Methods: Data for 50,996 patients diagnosed with non-metastatic colon cancer were retrieved from the Surveillance, Epidemiology, and End Results (SEER) database. Patients were assigned randomly to the training set ($n=34,168$) or validation set ($n=16,828$). Independent prognostic factors were identified by multivariate Cox proportional hazards regression analysis and used to construct the nomogram. Harrell's C-index and calibration plots were calculated using the SEER validation set. Additional external validation was performed using a Chinese dataset ($n=342$).

Results: Harrell's C-index of the nomogram for OS in the SEER validation set was 0.71, which was superior to that using the 7th edition of the American Joint Committee on Cancer TNM staging (0.59). Calibration plots showed consistency between actual observations and predicted 1-, 3-, and 5-year survival. Harrell's C-index (0.72) and calibration plot showed excellent predictive accuracy in the external validation set.

Conclusions: We developed a nomogram to predict OS after curative resection for colon cancer. Validation using the SEER and external datasets revealed good discrimination and calibration. This nomogram may help predict individual survival in patients with colon cancer.

¹Department of Gastrointestinal Surgery, The First Affiliated Hospital of Bengbu Medical College, Bengbu, Anhui, China

²Department of General Surgery, The Fourth Affiliated Hospital of Anhui Medical University, Hefei, Anhui, China

*These authors contributed equally to this work

Corresponding author:

Wenbin Wang, Department of General Surgery, The Fourth Affiliated Hospital of Anhui Medical University, 100 Huaihai Road, Hefei, Anhui Province 230000, China.
Email: wbwangay@163.com



Keywords

Colon cancer, nomogram, overall survival, SEER, prognosis, prediction

Date received: 11 December 2020; accepted: 13 April 2021

Introduction

Colon cancer is the fourth most common cancer and the fifth-leading cause of cancer-related deaths worldwide.¹ Improvements in surgical techniques and the application of comprehensive treatments have improved the long-term survival of patients with colon cancer.^{2,3} However, the survival of patients after radical colon cancer surgery varies greatly among different tumor, nodes, metastasis (TNM) classifications.

The 7th edition of the American Joint Committee on Cancer (AJCC) TNM classification released in 2010 classified M0 colon cancer into eight groups based on the depth of invasion and the number of metastatic lymph nodes.⁴ However, although this staging system is widely used to predict prognosis, patients with the same stage often have different clinical prognoses, suggesting that survival is also affected by other factors. Previous studies accordingly found that clinicopathological features such as sex, tumor site, and serum biomarkers affected the survival of patients with colon cancer.⁵⁻⁷

A nomogram is a predictive tool that creates a simple graphical representation of a statistical predictive model to generate the numerical probability of a clinical event.⁸ Nomograms have been widely applied to many types of tumors, including colon cancer.⁹⁻¹¹ However, few nomograms predicting the survival probability of patients with colon cancer have been validated in different populations.¹² The current study therefore aimed to develop and validate a nomogram for colon cancer based on the Surveillance, Epidemiology, and End Results (SEER) database, which contains data for both western and eastern patients,

and to additionally validate the nomogram using an external Chinese cohort.

Patients and methods

Patient selection

In this study, we extracted data from the SEER database using SEER*Stat, version 8.3.6 software (released 8 August 2019; <http://seer.cancer.gov/data/options.html>).

We identified patients with colon cancer diagnosed between 2010 and 2015. The inclusion criteria were: (1) colon cancer stage I–III; (2) primary site code (C18.0, C18.2, C18.3, C18.4, C18.5, C18.6, and C18.7); (3) adenocarcinoma (histology codes 8140–8147, 8210–8211, 8220–8221, 8260–8263, 8480–8481, and 8490); and (4) histologically confirmed diagnosis. The exclusion criteria were (1) cases with missing information; (2) no cancer-directed surgery; (3) radiation; and (4) death or no follow-up within 30 days. The included patients were divided randomly into a training set and a validation set (ratio 2:1). The results were also validated using data for patients diagnosed at the First Affiliated Hospital of Bengbu Medical College (BBMC) between 2015 and 2019. A flow chart of patient selection is shown in Figure 1.

This study was conducted in accordance with the World Medical Association's Declaration of Helsinki and approved by the Ethics Committee of Bengbu Medical College (approval number 2020-238). The study did not include any personally identifiable information, and the need for informed consent was therefore waived by our ethics committee.

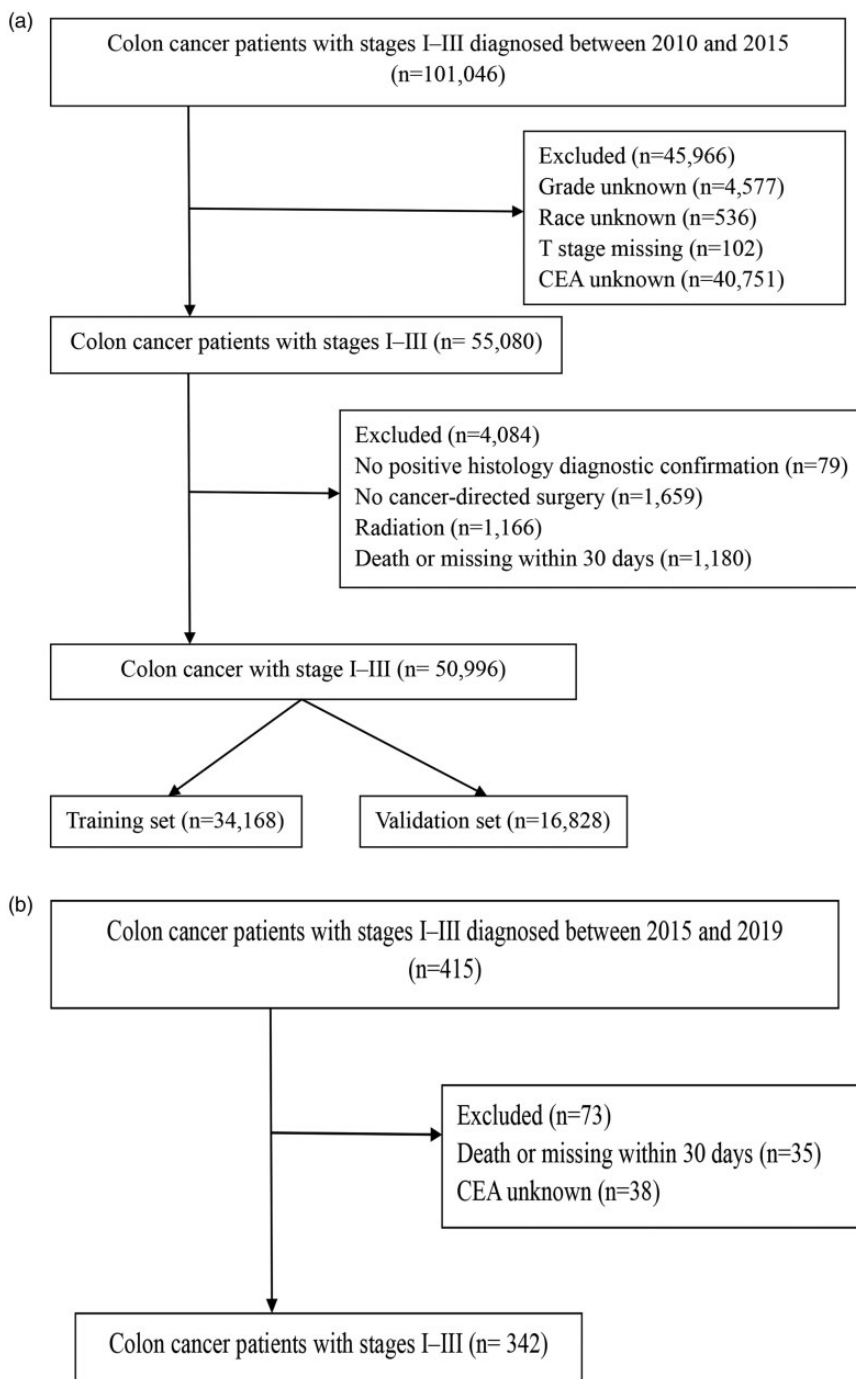


Figure 1. Flow chart of patient selection. (a) Surveillance, Epidemiology, and End Results database. (b) Bengbu Medical College database. CEA, carcinoembryonic antigen.

Variables

We extracted data for 10 clinicopathological variables, including sex, age, race, primary site, grade, TNM stage, T stage, N stage, adjuvant chemotherapy, and carcinoembryonic antigen (CEA) level. Tumor location was categorized as right-sided colon cancer (RCC) or left-sided colon cancer (LCC). RCCs included cancers of the cecum, ascending colon, hepatic flexure of colon, and transverse colon, and LCCs included cancers of the splenic flexure of the colon, descending colon, and sigmoid colon. Race was categorized as Asian or Pacific Islander (API) or non-API (white and black). Tumor staging was performed in accordance with the 7th edition of the AJCC TNM classification. Age was converted to a categorical variable. The primary endpoint was overall survival (OS). Patients in the SEER database were followed up from diagnosis until death or 31 December 2016. Patients in the BBMC database were followed up from diagnosis until death or 30 June 2020.

Construction of the nomogram

OS was estimated using the Kaplan–Meier method and validated by log–rank test. All significant variables in the univariate analysis were entered into multivariate Cox proportional hazards analysis. All the included variables satisfied the proportional hazards assumption based on a log–log survival plot. Prognostic factors were identified by a forward stepwise Cox proportional hazards regression model. A nomogram based on the results of this model was established to predict 1-, 3-, and 5-year OS.

Validation of the nomogram

The nomogram was validated by discrimination and calibration. Discrimination was evaluated by Harrell's concordance index (Harrell's C-index), which can deal with

censored data and estimate the probability of an event, with a higher C-index indicating better discrimination. The difference between predicted and actual survival was compared using calibration plots. The nomogram was validated using an internal SEER validation set ($n=16,828$) and an external BBMC validation set ($n=342$).

Statistical analyses

Statistical analysis was carried out using IBM SPSS Statistics for Windows, version 23.0 (IBM Corp., Armonk, NY, USA). The nomogram was constructed and verified using R software (version 3.6.1).¹³ A P -value <0.05 was considered statistically significant, and all statistical tests were two-sided.

Results

A total of 50,996 patients from the SEER database were included in the present study. There were 34,168 patients in the training set, 16,828 patients in the SEER validation set, and 342 patients in the BBMC validation set. The clinicopathological characteristics of the patients are shown in Table 1.

The potential variables in the training set were selected by forward stepwise selection. Univariate Cox regression analysis showed that age, race, primary site, grade, T stage, N stage, chemotherapy, and CEA level were associated with OS, and multivariate Cox regression analysis confirmed these eight variables as independent predictors of OS (Table 2). A nomogram predicting 1-, 3-, and 5-year OS was constructed based on these independent prognostic factors (Figure 2). Adding the points for each variable produced total scores predicting the probabilities of 1-, 3-, and 5-year OS.

The nomogram was validated using the SEER and BBMC validation sets. The clinicopathological characteristics of the patients in the BBMC validation set are

Table 1. Demographic and clinicopathologic characteristics of included patients.

Variable	Training set (n = 34,168)		SEER validation Set (n = 16,828)		BBMC validation Set (n = 342)	
	n	%	n	%	n	%
Sex						
Female	17,319	50.7	8486	50.4	152	44.4
Male	16,849	49.3	8342	49.6	190	55.6
Age (years)						
<65	12,906	37.8	6358	37.8	187	54.7
≥65	21,262	62.2	10,470	62.2	155	45.3
Race						
Non-API	31,009	90.8	15,210	90.4	0	0
API	3159	9.2	1618	9.6	342	100
Primary site						
RCC	21,698	63.5	10,691	63.5	181	52.9
LCC	12,470	36.5	6137	36.5	161	47.1
Grade						
Grade I	2731	8.0	1371	8.1	5	1.5
Grade II	24,992	73.1	12,373	73.5	223	65.2
Grade III	5390	15.8	2613	15.5	38	11.1
Grade IV	1055	3.1	471	2.8	76	22.2
TNM stage						
Stage I	8333	24.4	4230	25.1	35	10.2
Stage II	12,802	37.5	6306	37.5	188	55.0
Stage III	13,033	38.1	6292	37.4	119	34.8
T stage						
T1	4342	12.7	2212	13.1	9	2.6
T2	5653	16.5	2821	16.8	39	11.4
T3	19,448	56.9	9475	56.3	0	0
T4	4725	13.8	2320	13.8	294	86.0
N stage						
N0	21,139	61.9	10,536	62.6	223	65.2
N1	8660	25.3	4173	24.8	82	24.0
N2	4369	12.8	2119	12.6	37	10.8
Chemotherapy						
No	23,375	68.4	11,607	69.0	79	23.1
Yes	10,793	31.6	5221	31.0	263	76.9
CEA						
Negative	22,085	64.6	10,922	64.9	213	62.3
Positive	12,083	35.4	5906	35.1	129	37.7

SEER, Surveillance, Epidemiology, and End Results; BBMC, Bengbu Medical College; CEA, carcinoembryonic antigen; API, Asian or Pacific Islander.

shown in Table 1. Harrell's C-index for the SEER validation set was 0.71 (95% confidence interval (CI), 0.65–0.77), which was superior to that for the seventh edition of the AJCC TNM stage (0.59; 95%CI, 0.55–

0.63) ($P < 0.001$). Calibration plots of the SEER validation of OS showed satisfactory consistency between the actual observed and predicted probabilities (Figure 3). Harrell's C-index for the BBMC validation

Table 2. Univariate and multivariate analyses of overall survival in the training set.

Variable	Univariate analysis			Multivariate analysis		
	HR	95%CI	P	HR	95%CI	P
Sex			0.320			
Female	1					
Male	1.022	0.979–1.022				
Age (years)			<0.001			<0.001
<65	1			1		
≥65	2.487	2.361–2.619		2.212	2.096–2.335	
Race			<0.001			<0.001
Non-API	1			1		
API	0.828	0.765–0.897		0.807	0.745–0.874	
Primary site			<0.001			<0.001
RCC	1			1		
LCC	0.731	0.698–0.766		0.873	0.833–0.916	
Grade			<0.001			<0.001
Grade I	1			1		
Grade II	1.211	1.110–1.322	<0.001	1.036	0.949–1.133	0.428
Grade III	1.890	1.718–2.079	<0.001	1.234	1.118–1.361	<0.001
Grade IV	2.274	1.996–2.589	<0.001	1.395	1.222–1.593	<0.001
T stage			<0.001			<0.001
T1	1			1		
T2	1.222	1.108–1.349	<0.001	1.050	0.951–1.160	0.333
T3	1.780	1.640–1.933	<0.001	1.342	1.232–1.462	<0.001
T4	3.602	3.296–3.937	<0.001	2.501	2.273–2.752	<0.001
N stage			<0.001			<0.001
N0	1			1		
N1	1.436	1.365–1.509	<0.001	1.874	1.771–1.983	<0.001
N2	2.483	2.351–2.623	<0.001	3.034	2.845–3.235	<0.001
Chemotherapy			<0.001			<0.001
No	1			1		
Yes	0.797	0.760–0.836		0.473	0.447–0.500	
CEA			<0.001			<0.001
Negative	1			1		
Positive	1.866	1.789–1.947		1.526	1.461–1.594	

HR, hazard ratio; CEA, carcinoembryonic antigen; API, Asian or Pacific Islander.

set was 0.72 (95%CI, 0.65–0.79). The calibration plot of the nomogram indicated no deviations from the reference line (Figure 4).

Discussion

In this study, we developed a nomogram based on the SEER database to predict OS in patients with resectable colon

cancer. We also validated this prognostic model in an independent external cohort. The nomogram showed better predictive accuracy than the traditional TNM staging system in patients with resectable colon cancer. The significant clinicopathological variables included in this novel nomogram were age, race, primary site, grade, chemotherapy, and CEA level.

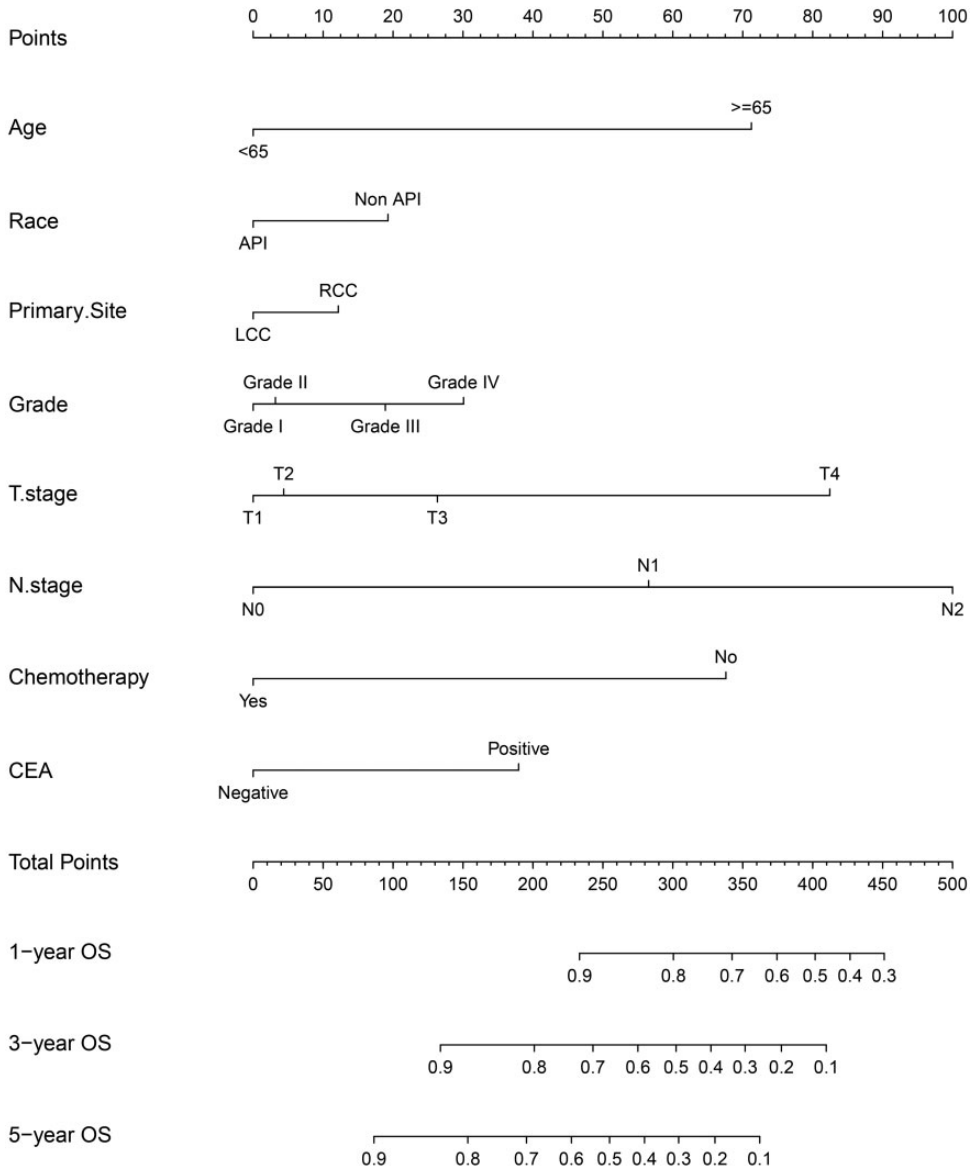


Figure 2. Nomogram predicting 1-, 3-, and 5-year overall survival (OS) of patients with colon cancer after curative resection.

The current nomogram had certain advantages and key features compared with previous nomograms for resectable colon cancer. First, the present population was not age-restricted, compared with some studies that only included elderly patients¹⁴

or younger patients.¹¹ In addition, some previous studies only performed internal validation,¹⁵ while we validated the nomogram using both internal and external datasets, given that validation using a dataset from other countries is generally considered

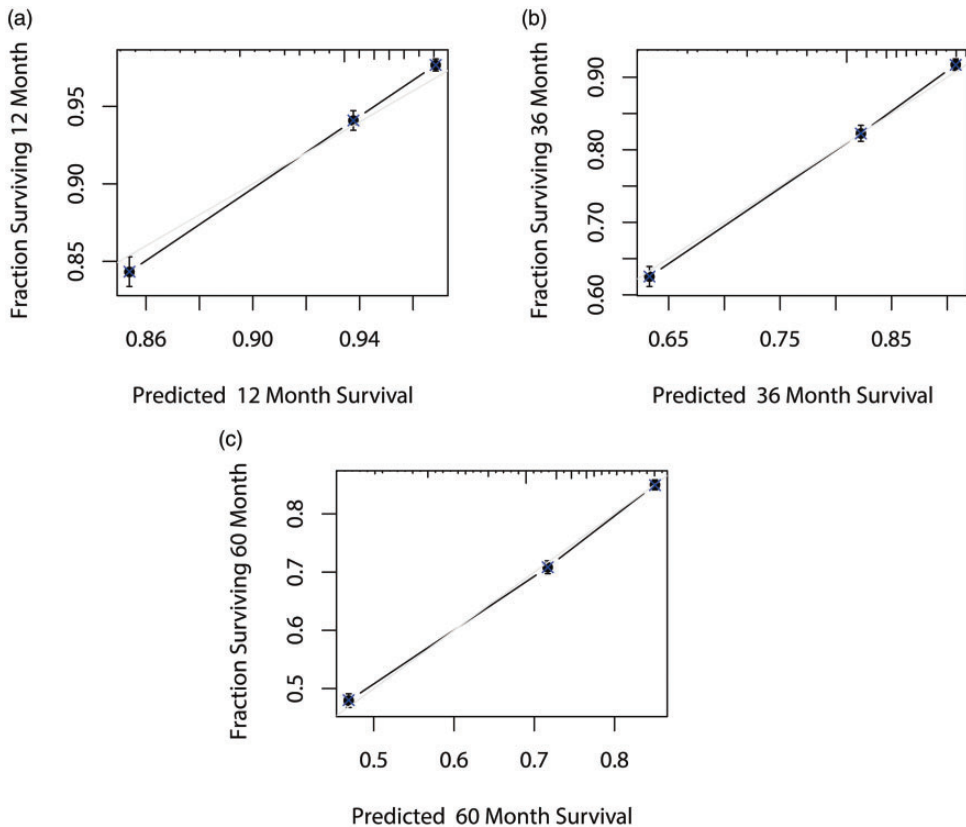


Figure 3. Calibration plots of the Surveillance, Epidemiology, and End Results validation set. X-axes represent nomogram-predicted probability of survival and y-axes represent actual survival. The reference line is gray. (a) 1-year, (b) 3-year, and (c) 5-year calibration plots.

as the most stringent form of external validation.¹⁶

Several previous studies found associations between sex and survival. One study found that female sex was associated with a poor prognosis in patients with stage III colon cancer,⁵ while another showed that female sex was associated with higher lymph node yield, which implied an increase in OS in stage I–III colorectal cancer.¹⁷ However, no previous studies have reported on the relationship between female sex and survival in patients with stage I–III colon cancer, and the current study found that sex was not an independent prognostic factor in patients with stage I–III colon cancer.

Elderly patients with stage I–III colon cancer have a higher risk of death than younger patients. Many studies have confirmed that postoperative mortality increases with increasing age.^{18,19} Elderly patients tend to have more comorbidities than younger patients, especially cardiovascular and respiratory diseases, and often also have large and locally invasive tumors.²⁰ In addition, although adjuvant chemotherapy can improve the long-term survival of patients with stage III colon cancer, half of all elderly patients with stage III colon cancer do not receive adjuvant chemotherapy²¹ and have a higher risk of death. It is also possible that younger

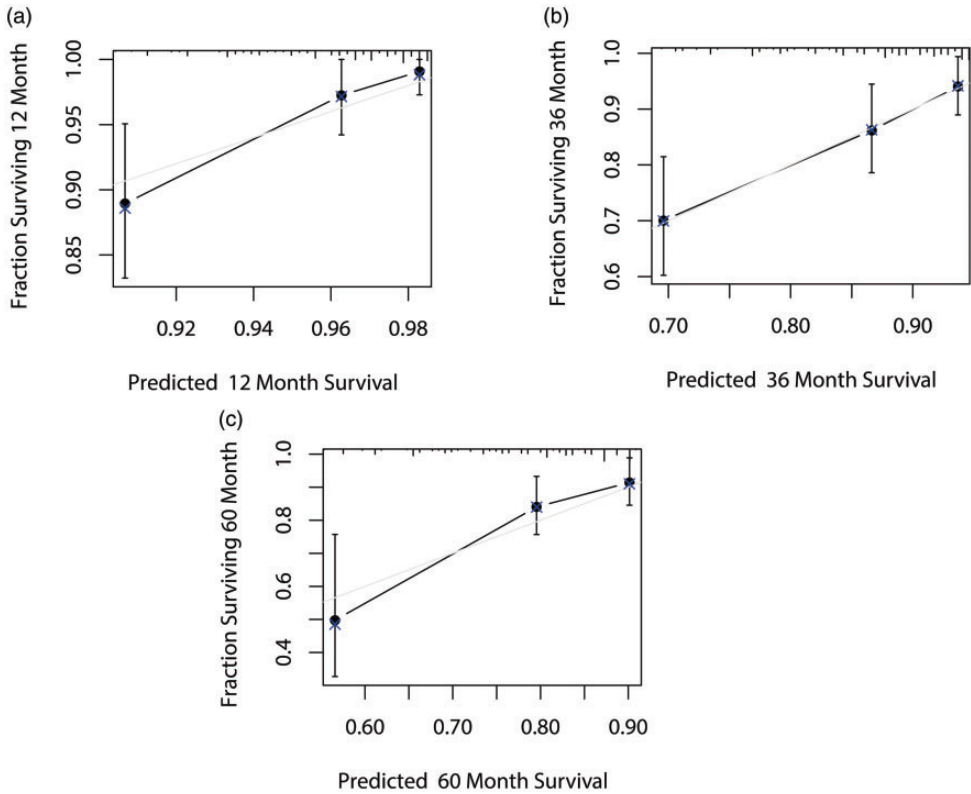


Figure 4. Calibration plots of the Bengbu Medical College validation set. X-axes represent nomogram-predicted probability of survival and y-axes represent actual survival. The reference line is gray. (a) 1-year, (b) 3-year, and (c) 5-year calibration plots.

patients may derive greater OS benefit from adjuvant chemotherapy compared with elderly patients.²²

The present study identified race as an independent prognostic factor in patients with colon cancer, with API patients having a better prognosis than non-API patients. Similar results have been reported in previous studies of patients with colon cancer and gastric cancer.^{23,24} Histological grade was also confirmed as an adverse prognostic factor.²⁵ Patient prognosis could thus be predicted more precisely by incorporating race and histological grade into the nomogram.

In this study, we categorized the tumor location as LCC or RCC. RCC has been

associated with a significantly increased risk of death, independent of stage, race, and adjuvant chemotherapy^{6,26}; CEA has also been identified as a prognostic factor for all stages of colon cancer, with patients with an elevated CEA level having a 62% increase in the risk of death (hazard ratio 1.62, 95%CI 1.53–1.74) compared with patients with a normal CEA level.²⁷ As shown by the current nomogram, patients with RCC and high preoperative CEA tended to have a significantly poorer prognosis.

Accurate prognostic assessment is critical for both doctors and patients. The current nomogram was more precise than the 7th edition of the AJCC TNM staging

system, allowing the development of individualized follow-up strategies. However, although our nomogram was verified externally and showed good accuracy, there were several limitations. First, although we knew whether or not patients in the SEER database had received chemotherapy, we did not know the specific scheme of chemotherapy, and different schemes may lead to bias. Second, our nomogram was based on the SEER database, and all the analyses used data available in that database. However, information on some known survival predictors, such as *BRAF*, *KRAS*, and *TP53* mutations, CpG island methylator phenotype, mismatch repair status, and chromosomal instability status, was not included.²⁸ Comorbidity, performance status, socioeconomic status, smoking, and other lifestyle factors may also influence survival, and these data were also lacking in the SEER database. Finally, although we carried out external verification, the number of patients in the external cohort was small, and the model needs to be further verified using a larger sample size.

Conclusions

We developed a nomogram to predict 1-, 3-, and 5-year OS in patients with colon cancer following curative resection. Validation of the nomogram using the SEER and BBMC datasets revealed good discrimination and calibration. This nomogram may thus be helpful for predicting individual survival among patients with resected colon cancer.

Acknowledgement

We thank LetPub (www.letpub.com) for linguistic assistance during the preparation of this manuscript.

Declaration of conflicting interest


The authors declare that there is no conflict of interest.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Innovation Team Project of Bengbu Medical College [grant number BYKC201909].

ORCID iDs

Shuanhu Wang  <https://orcid.org/0000-0002-5439-3482>

Mulin Liu  <https://orcid.org/0000-0002-8655-6207>

References

1. Bray F, Ferlay J, Soerjomataram I, et al. 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. DOI: 10.3322/caac.21492.
2. Andre T, Boni C, Mounedji-Boudiaf L, et al. Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer. *N Engl J Med* 2004; 350: 2343–2351. DOI: 10.1056/NEJMoa032709.
3. White A, Joseph D, Rim SH, et al. Colon cancer survival in the United States by race and stage (2001-2009): Findings from the CONCORD-2 study. *Cancer* 2017; 123: 5014–5036. DOI: 10.1002/cncr.31076.
4. Balch CM, Gershenwald JE, Soong SJ, et al. Final version of 2009 AJCC melanoma staging and classification. *J Clin Oncol* 2009; 27: 6199–6206. DOI: 10.1200/JCO.2009.23.4799.
5. Park JH, Park HC, Park SC, et al. Female sex and right-sided tumor location are poor prognostic factors for patients with stage III colon cancer after a curative resection. *Ann Coloproctol* 2018; 34: 286–291. DOI: 10.3393/ac.2018.10.29.
6. Yahagi M, Okabayashi K, Hasegawa H, et al. The worse prognosis of right-sided compared with left-sided colon cancers: a systematic review and meta-analysis. *J Gastrointest Surg* 2016; 20: 648–655. DOI: 10.1007/s11605-015-3026-6.
7. Kim CG, Ahn JB, Jung M, et al. Preoperative serum carcinoembryonic

- antigen level as a prognostic factor for recurrence and survival after curative resection followed by adjuvant chemotherapy in stage III colon cancer. *Ann Surg Oncol* 2017; 24: 227–235. DOI: 10.1245/s10434-016-5613-5.
8. Iasonos A, Schrag D, Raj GV, et al. How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol* 2008; 26: 1364–1370. DOI: 10.1200/JCO.2007.12.9791.
 9. Chen S, Chen S, Lian G, et al. Development and validation of a novel nomogram for pre-treatment prediction of liver metastasis in pancreatic cancer. *Cancer Med* 2020; 9: 2971–2980. DOI: 10.1002/cam4.2930.
 10. Han DS, Suh YS, Kong SH, et al. Nomogram predicting long-term survival after d2 gastrectomy for gastric cancer. *J Clin Oncol* 2012; 30: 3834–3840. DOI: 10.1200/JCO.2012.41.8343.
 11. Yu C and Zhang Y. Development and validation of a prognostic nomogram for early-onset colon cancer. *Biosci Rep* 2019; 39: BSR20181781. DOI: 10.1042/BSR20181781.
 12. Liu M, Qu H, Bu Z, et al. Validation of the Memorial Sloan-Kettering Cancer Center nomogram to predict overall survival after curative colectomy in a Chinese colon cancer population. *Ann Surg Oncol* 2015; 22: 3881–3887. DOI: 10.1245/s10434-015-4495-2.
 13. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2020. <https://www.R-project.org/>
 14. 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: 202–208. DOI: 10.1080/00365521.2020.1720280.
 15. Zhou Z, Mo S, Dai W, et al. Prognostic nomograms for predicting cause-specific survival and overall survival of stage I-III colon cancer patients: a large population-based study. *Cancer Cell Int* 2019; 19: 355. DOI: 10.1186/s12935-019-1079-4.
 16. Harrell FE Jr, Lee KL and Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996; 15: 361–387. DOI: 10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4.
 17. Ahmadi O, Stringer MD, Black MA, et al. Clinico-pathological factors influencing lymph node yield in colorectal cancer and impact on survival: analysis of New Zealand Cancer Registry data. *J Surg Oncol* 2015; 111: 451–458. DOI: 10.1002/jso.23848.
 18. Bos A, Kortbeek D, Van Erning FN, et al. Postoperative mortality in elderly patients with colorectal cancer: The impact of age, time-trends and competing risks of dying. *Eur J Surg Oncol* 2019; 45: 1575–1583. DOI: 10.1016/j.ejso.2019.04.020.
 19. Hermans E, Van Schaik PM, Prins HA, et al. Outcome of colonic surgery in elderly patients with colon cancer. *J Oncol* 2010; 2010: 865908. DOI: 10.1155/2010/865908.
 20. Itatani Y, Kawada K and Sakai Y. Treatment of elderly patients with colorectal cancer. *Biomed Res Int* 2018; 2018: 2176056. DOI: 10.1155/2018/2176056.
 21. Merchant SJ, Nanji S, Brennan K, et al. Management of stage III colon cancer in the elderly: Practice patterns and outcomes in the general population. *Cancer* 2017; 123: 2840–2849. DOI: 10.1002/cncr.30691.
 22. Meyers BM, Cosby R, Quereshy F, et al. Adjuvant chemotherapy for stage II and III colon cancer following complete resection: A cancer care Ontario Systematic Review. *Clin Oncol (R Coll Radiol)* 2017; 29: 459–465. DOI: 10.1016/j.clon.2017.03.001.
 23. Hashiguchi Y, Hase K, Ueno H, et al. Impact of race/ethnicity on prognosis in patients who underwent surgery for colon cancer: analysis for white, African, and East Asian Americans. *Ann Surg Oncol* 2012; 19: 1517–1528. DOI: 10.1245/s10434-011-2113-5.
 24. Theuer CP, Kurosaki T, Ziogas A, et al. Asian patients with gastric carcinoma in the United States exhibit unique clinical features and superior overall and cancer specific survival rates. *Cancer* 2000; 89: 1883–1892. DOI: 10.1002/1097-0142(20001101)89:9<1883::aid-cnrc3>3.3.co;2-8.

25. Nanni O, Volpi A, Frassinetti GL, et al. Role of biological markers in the clinical outcome of colon cancer. *Br J Cancer* 2002; 87: 868–875. DOI: 10.1038/sj.bjc.6600569.
26. Petrelli F, Tomasello G, Borgonovo K, et al. Prognostic survival associated with left-sided vs right-sided colon cancer: a systematic review and meta-analysis. *JAMA Oncol* 2017; 3: 211–219. DOI: 10.1001/jamaoncol.2016.4227.
27. Becerra AZ, Probst CP, Tejani MA, et al. Evaluating the prognostic role of elevated preoperative carcinoembryonic antigen levels in colon cancer patients: results from the National Cancer Database. *Ann Surg Oncol* 2016; 23: 1554–1561. DOI: 10.1245/s10434-015-5014-1.
28. Marisa L, De Reynies A, Duval A, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med* 2013; 10: e1001453. DOI: 10.1371/journal.pmed.1001453.