Analysis of Clinical Characteristics and Prognosis of Patients With Stages I-III Colon Cancer Based on Seer Data

Fuqiang Zhao (✉ 247219491@qmu.edu.cn)

Qiqihaer Medical College: Qiqihar Medical University

Ying Sun

Qiqihaer Medical College: Qiqihar Medical University

Jie Ge

Qiqihaer Medical College: Qiqihar Medical University

Chunlei Zheng

Qiqihaer Medical College: Qiqihar Medical University

Kepeng Ning

Qiqihaer Medical College: Qiqihar Medical University

Jingying Zhao

Qiqihaer Medical College: Qiqihar Medical University

Fangxu Zhu

Qiqihaer Medical College: Qiqihar Medical University

Research Article

Keywords: SEER database, colon cancer, prognosis, nomogram

Posted Date: July 16th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-708820/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Objective: This study analyzes the clinical features and prognosis of stages I-II colon cancer, and constructs a nomogram to predict the prognosis of patients.

Methods: We included data from patients with stages I-II colon cancer confirmed by pathology after surgical treatment in the United States SEER (Surveillance, Epidemiology, and End Results) database from 2010-2015. The included patients are randomly divided into training cohorts and validation cohorts (ratio 1:1). The independent related factors of the prognosis of colon cancer patients were used to construct the nomogram and the web-based probability calculator. Concordance index (C-index) and calibration curve are used to evaluate the accuracy of the model.

Results: After univariate and multivariate analysis, it was indicated that age, marital status, tumor grade, t-stage and n-stage were independent factors affecting prognosis of patients with stages I-II colon cancer. We built a nomogram and the web-based probability calculator based on this, and its C-index was 0.781 (95% CI: 0.77414–0.78786), and were superior to that of AJCC TNM Stage (C-index: 0.734, 95%CI: 0.72616–0.74184). The consistency test showed that the nomogram can effectively predict the prognosis of patients. The validation cohort confirms the reliability of the model.

Conclusion: Age, marital status, tumor grade, t-stage and n-stage are independent factors affecting the prognosis of patients with stages I-II colon cancer. The nomogram we constructed in this way can better predict the prognosis of patients with stages I-II colon cancer.

Introduction

According to the statistics of the World Health Organization, the incidence of colorectal cancer in the world in 2018 is 1849518 cases, accounting for the third place among all malignant tumors. There were 880792 deaths due to colorectal cancer, accounting for the second place among all malignant tumors [1]. Although patients with stages I-II colon cancer have a higher five-year survival rate, the base of patients with stages I-II colon cancer is larger, and a considerable number of patients still have local recurrence or distant metastasis after surgery. It is found that patients with colon cancer in the same TNM stage have different prognosis and response to treatment, which indicates that there are other factors that affect the progression of colon cancer. The TNM staging system only considers the anatomical characteristics of the tumor, but did not consider the biological characteristics of the tumor. In Clinically, variables related to tumor growth characteristics need to be incorporated into individualized treatment. Nomogram is widely used for malignancy Tumor prognosis, it is a simple and practical prognostic model that can combine multiple clinically valuable prognostic factors to establish a prognostic model and simplify to assess the probability of an individual patient's event (such as death or recurrence).

The SEER database is a population-based tumor epidemiology database in the United States, covers approximately 34.6% of the population in the United States, containing 18 registry cases since 1973 and detailed clinical and prognostic information, research on colon cancer and other tumors with great

help^[2,3]. This study retrospectively analyzed the clinicopathological data of patients with stages I-III colon cancer undergoing surgical treatment, and used a COX proportional hazards regression model in survival analysis to screen for multiple independent risk factors affecting stages I-III colon cancer patients and construct the nomogram prognosis prediction model.

Methods

1.1 Data sources

We downloaded the SEER database (<http://seer.cancer.gov/>) through the SEER * Stat software (v8.3.6.0, <https://seer.cancer.gov/seerstat/>). All patients with primary stages I-III colon cancer undergoing surgery from 2010 to 2015. The inclusion criteria were: primary colon cancer, American Joint Committee on Cancer (AJCC) stage is the stages I-III of colon cancer, and pathologically confirmed adenocarcinoma.

The exclusion criteria were: The cause of death is unclear(207); Patients with a survival time of 0 (3447). The eligible patients were randomly divided into a training cohort and a validation cohort (ratio 1:1, Table1). We extracted and analyzed variables such as age, race, sex, marital status, primary tumor site, degree of tumor differentiation, number of lymph node, TNM stage, surgery, and survival status and time. In order to reduce the impact of other factors on the prognosis of stages I-III colon cancer, the end point of this study was cancer-specific death.

1.2 Statistical analysis

Use SPSS for Windows 22.0 statistical software package for statistical analysis. Cases were randomly divided into training cohorts and validation cohorts (ratio 1:1). Descriptive statistical methods were used to summarize the total case distribution. The count data is expressed as a percentage, and the chisquare test was applied to compare the composition ratio of the training and the validation cohorts. The Kaplan-Meier method was used to calculate the cancer specific survival (CSS). Use log-rank test to compare differences in survival rates. In the univariate factor analysis, the variables with $P < 0.05$ were analyzed by the COX multivariate regression model. The significance of learning is a two-sided test.

Import independent prognostic factors into R4.0.2 software, and use the "rms" and "DynNom" software package to establish a nomogram prediction model and the web-based probability calculator. The model is a visual process of COX multi-factor analysis. Each relevant factor corresponds to the scale in the figure. The scores of all factors are added to obtain a total score, and cancer specific survival (CSS) of 1-year, 3-year and 5-year are obtained from the total score. The Bootstrap method is used to calculate the c-index, draw a calibration chart of predicted and actual values, and evaluate the predicted model. And compared with the eighth edition of the international malignant tumor TNM staging. The accuracy of the nomogram model is measured by c-index. When the c-index is closer to 1, it indicates that the accuracy of the prediction model is higher. The consistency check of the model is achieved by drawing a calibration chart of the predicted and true values of the model.

Results

2.1 Clinical features of stages I-II colon cancer

A total of 73514 patients with stages I-II colon adenocarcinoma confirmed by surgical pathology based on the inclusion and exclusion criteria of SEER database were enrolled. Among all patients, 50.64% were male, 16.43% were single, 78.30% were white and 60.87% were proximal colon. At the end of follow-up, a total of 7707 (10.48%) patients with stages I-II colon adenocarcinoma died of cancer-specific causes.

2.2 Analysis of prognostic factors for stages I-II colon cancer

We used univariate factor analysis to investigate the possible prognostic factors of patients with stages I-II colon cancer and found that the correlation was statistically significant between age, sex, tumor site, tumor grade, t-stage, n-stage and marital status and the cancer specific survival of patients with stages I-II colon cancer (All $P < 0.05$). Further incorporating these statistically significant variables in univariate analysis into multivariate analysis revealed that age, tumor grade, t-stage, n-stage and marital status are independent prognostic factors that affect the specific survival of stages I-II colon cancer (All $P < 0.05$). (Figure 1).

2.3 Production and inspection of nomogram

We successfully constructed a nomogram and the web-based probability calculator (<https://coloncancernomogram.shinyapps.io/DynNomapp/>) based on the above independent predictors of patient prognosis (Figure 2). Based on the patient's age, tumor grade, t-stage, n-stage and marital status, we can calculate the patient's Cancer specific survival rates at 1-year, 3-year and 5-year. After the discrimination test, the C-index of this nomogram is 0.781 (95% CI: 0.77414–0.78786), which is higher than the eighth edition of the TNM staging (C-index: 0.734, 95%CI: 0.72616–0.74184). The consistency test shows that the predicted 1-year, 3-year and 5-year cancer specific survival rates of the nomogram are in good agreement with the actual 1-year, 3-year and 5-year cancer specific survival rates, and the slope of the consistency curve is close to 1. (Figure 3). To prove whether the nomogram can distinguish patients with different risks, the nomogram score of each patient was calculated, and all patients were divided into high risk group and low risk group according to the score. The Kaplan–Meier curves showed that the survival of the two groups was present statistically significantly different, indicating that the nomogram can distinguish patients with different risks (Figure 4).

Discussion

Despite significant progress in the diagnosis and treatment of disease, Colon cancer remains one of the leading causes of cancer-related mortality worldwide^[4]. At present, TNM staging is considered to be the gold standard for predicting the prognosis of colon cancer. Colon cancer showed recurrence diversity and survival heterogeneity at the same disease stage. Obtaining accurate prognostic information is essential to help doctors make better clinical decisions and consult on life expectancy after surgical removal of

tumors. Nomogram is widely used to predict the prognosis of various cancers, and compared with the traditional TNM staging system, the nomogram is a simpler tool with many advantages. Personalized risk prediction for patients is one of the main advantages of nomograms^[5, 6]. Some studies on the prognosis of colon cancer have been reported. However, the number of patients with stages I-II colon cancer in these studies is small^[7,8].

In this study, We conducted an in-depth analysis of the data of patients with stages I-II colon cancer after tumor resection in the United States SEER database. We found that age, tumor grade, t-stage, n-stage and marital status are independent factors affecting the prognosis of patients with stages I-II colon cancer. These variables were used to draw a nomogram. The nomogram we constructed in this way can better predict the specific 1-year, 3-year and 5-year cancer specific survival rates of patients. Consistency tests show that the nomogram can effectively predict the specific 1-year, 3-year and 5-year cancer specific survival rates of patients. The nomogram C- index of cancer specific survival rate is 0.781 (95% CI: 0.77414–0.78786), which is better than AJCC TNM stage (C-index: 0.734, 95%CI: 0.72616–0.74184). Based on the research results already obtained, the nomogram clearly shows better discrimination than the TNM staging system.

This is the first nomogram study among patients with stages I-II colon cancer in the SEER database. Complications are common in patients with colon cancer and can affect clinical decisions and survival prognosis. Therefore, the end point event selected in this study was cancer specific survival, which minimized the impact of comorbidities on survival and more accurately highlighted the impact of colon cancer on its prognosis. This is a feasible strategy for the treatment of patients with stages I-II colon cancer and can predict their individual survival prognosis.

By analyzing cases in the entire population in the SEER database, you can effectively avoid the bias of patients from a single institution to the study. However, the SEER database lacks information such as imaging, smoking history, gene mutations, tumor markers, and detailed treatment methods. Our study also did not involve the impact of these factors on the prognosis of patients with stages I-II colon cancer, and these factors may seriously affect the prognosis of patients with stages I-II colon cancer. Second, it is impossible to obtain comorbidities and adjuvant treatment information (including radiotherapy and chemotherapy) for all patients. Deleting those incomplete cases may lead to selection bias, which may affect the prediction of the survival prognosis of patients with stages I-II colon cancer, Thereby reducing the number of patients counted and the prediction accuracy of the nomogram.

In summary, we have developed a valuable nomogram to estimate the specific survival probability of cancer in patients with stages I-II colon cancer at 1-year, 3-year and 5-year. The nomogram shows good performance in stages I-II colon cancer and can be used as valuable tool for clinical decision-making consultation and guidance for patients with stages I-II colon cancer.

Declarations

Funding:Science and Technology Research Project in Department of Education of Heilongjiang Provincial China. (2020-KYYWF-0032)

Conflicts of interest:The authors declare that they have no competing interests.

Ethics approval:We received permission to access the research data file in the SEER program from the National Cancer Institute. Approval was waived by the local ethics committee, as SEER data is publicly available and de-identified.

Consent to participate:We received permission to access the research data file in the SEER program from the National Cancer Institute. Approval was waived by the local ethics committee, as SEER data is publicly available and de-identified.IK

Consent for publication:No individually identifiable data are presented.

Availability of data and material:We received permission to access the research data file in the SEER program from the National Cancer Institute, United States.

Authors' contributions :Fuqiang Zhao and Jie Ge designed the research; Fuqiang Zhao, Jingying Zhao and Kepeng Ning analyzed data; Fuqiang Zhao, Fangxu Zhu, Chunlei Zheng wrote the draft. All members have proofread the article.

Acknowledgments The authors thank SEER program for public access to the database.

References

1. <https://gco.iarc.fr/today/fact-sheets-cancers>.
2. Yang X, Zhan C, Li M, et al. Lobectomy Versus Sublobectomy in Metachronous Second Primary Lung Cancer: A Propensity Score Study. *Ann Thorac Surg*. 2018;106(3):880-887. doi:10.1016/j.athoracsur.2018.04.071.
3. Yang X, Sun F, Chen L, et al. Prognostic value of visceral pleural invasion in non-small cell lung cancer: A propensity score matching study based on the SEER registry. *J Surg Oncol*. 2017;116(3):398-406. doi:10.1002/jso.24677.
4. Gasser M, Gerstlauer C, Grimm M, et al. Comparative analysis of predictive biomarkers for therapeutical strategies in colorectal cancer. *Ann Surg Oncol*. 2007;14(4):1272-1284. doi:10.1245/s10434-006-9155-0.
5. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol*. 2008;26(8):1364-1370. doi:10.1200/JCO.2007.12.9791.
6. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. *Lancet Oncol*. 2015;16(4):e173-e180. doi:10.1016/S1470-2045(14)71116-7.

7. Surgery for colorectal cancer in elderly patients: a systematic review. Colorectal Cancer Collaborative Group. *Lancet*. 2000;356(9234):968-974.

8. Hermans E, van Schaik PM, Prins HA, Ernst MF, Dautzenberg PJ, Bosscha K. Outcome of colonic surgery in elderly patients with colon cancer. *J Oncol*. 2010;2010:865908. doi:10.1155/2010/865908.

Tables

Table 1. Baseline characteristics of the patients included in the study.

	Total N=73514 No. (%)	Training cohort N=36757 No.(%)	Validation cohort N=36757 No. (%)	P-value
Race				.240
White	57562 78.30%	2876378.25%	2879978.35%	
Black	889712.10%	441312.01%	448412.20%	
Other	66599.06%	33915.32%	32688.89%	
NA	396	190	206	
Sex				.155
Male	36285 49.36%	1823949.62%	1804649.10%	
Female	37229 50.64%	1851850.38%	1871150.90%	
Marital				.103
single	12078(16.43%)	6121(16.65%)	5957(16.21%)	
married	61436(83.57%)	30636(83.35%)	30800(83.79%)	
Site				.241
Proximal colon	44749 60.87%	2228560.63%	2246461.11%	
Distal colon	27499 37.41%	1381837.59%	1368137.33%	
NA	1266	654	612	
Grade				.770
GradeI	83799.55%	420411.44%	417511.36%	
GradeII	51702 69.48%	2586770.37%	2583570.29%	
GradeIII	11088 15.57%	550014.96%	558815.20%	
GradeIV	23453.51%	11863.23%	11593.15%	
T-stage				.928
Tis+T1	13497	673818.33%	675918.39%	

	18.36%		
T2	11221 15.26%	5641 15.355%	5580 15.18%
T3	38014 51.71%	18992 51.67%	19022 51.75%
T4	10458 14.23%	5215 14.19%	5243 14.26%
NA	324	171	153
N-stage			.550
N0	47329 74.23%	23611 64.24%	23718 64.53%
N1	17350 23.60%	8736 23.77%	8614 23.43%
N2	8551 11.63%	4261 11.59%	4290 11.67%
NA	284	149	135
nLN			.402
<12	13030 17.72%	6558 17.84%	6472 17.61%
≥12	60384 82.14%	30147 82.02%	30237 82.26%
NA	100	52	48

nLN:number of lymph nodes

Figures

Figure 1

Univariate **A** and multivariate **B** analysis forest plots.

Figure 2

According to the prediction model of the prognosis of stage II-III colon cancer, the score of this factor was obtained according to the corresponding scale of each factor, and the total score was obtained by adding the scores of each factor, and the corresponding 1-year, 3-year and 5-year cancer-specific survival rate of the patients could be obtained from the total score.

Figure 3

Calibration curve of the nomogram show the 1-year, 3-year and 5-year cancer-specific survival in the training cohort (A,B,C) and the validation cohort (D,E,F),The x-axes indicate the predicted survival calculated by the nomogram the y-axes are actual survival calculated by the Kaplan–Meier method.

Figure 4

Kaplan–Meier curves of cancer-specific survival according to the nomogram score in the training cohort (A) and the validation cohort (B).