# Competing-risks model for predicting the prognosis of stage II colon cancer

Fu-qiang Zhao[1] and Ying Sun[1]

[1]Affiliation not available

June 23, 2020

**Abstract**

Objectives: This study performed a competing-risks analysis using data from the SEER database on stage II colon cancer patients with the aim of identifying more accurate prognostic factors. Methods: Data on patients with stage II colon cancer were extracted from the SEER database. A univariate analysis used the cumulative incidence function and Gray's test, while multivariate analysis was performed using the Fine-Gray model. Cumulative hazards were compared with a competing-risks model constructed using KaplanMeier estimation. Results: The multivariate Fine-Gray analysis indicated that grade III/IV, stage T4, number of lymph nodes (nLN<12) were statistically significant. The results obtained using multivariate Cox regression were different, while Kaplan-Meier curve analysis led to an overestimation of the cumulative risk of stage II colon cancer patients. Conclusions: This study established a competing-risks analysis model for the first time based on the SEER database for the risk assessment of stage II colon cancer patients. The results may help clinicians to better understand stage II colon cancer and provide these patients with more appropriate support.

## 1 Introduction

Although there have been some reports of risk factors associated with stage II colon cancer survival, most of them were based on Cox proportional-hazards regression models and Kaplan-Meier estimates.Performing a follow-up or making observations of a two-state model from a start event to an event of interest is a common design and analysis strategy, and Kaplan-Meier estimates, log-rank tests, and Cox regression are widely used for such single events of interest. However, those statistical analyses involve only one type of event. In medical research, the observed endpoints are rarely single, but there are multiple endpoints. The occurrence of competing events "blocks" the occurrence of the ending event of interest and forms a "competing relationship." For example, in cardiovascular disease mortality studies, patients may die from cardiovascular disease or die from other causes such as cancer, suicide, etc The frequency of comorbidities may be especially high in older people; for example, the risk of death from heart disease and cerebrovascular disease increases with age in patients with nonsmall-cell lung cancer[1].Traditional survival analysis will treat such competing risks by censoring, which will lead to miscalculations of the survival function[2]. This is because the Kaplan-Meier method and the Cox method treat other competing events as censored, and there may be conclusions that are estimated to be high or even contrary with the facts, also called competitive risk bias[3]. These considerations indicate the need to use a competing-risks model to handle multiple endpoints. 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[4,5].This study conducted a competing-risks analysis using data from the SEER database on stage II colon cancer patients with the aim of identifying more accurate prognostic factors.

## 2 Methods and materials

### 2.1 patients

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 stage II colon cancer undergoing surgery from 2010 to 2015. The inclusion criteria were: primary colon cancer, American Joint Committee on Cancer (AJCC) stage is the stage II of colon cancer, and pathologically confirmed adenocarcinoma.

The exclusion criteria were: colon cancer with unknown primary site and overlapping lession of colon; Patients older than 85 years old; Dead(missing/unknow COD) (Figure 1). We extracted and analyzed variables such as age, race, sex, primary tumor site, degree of tumor differentiation, number of lymph node, TNM stage, surgery, and survival status and time.

### 2.2 Statistical analysis

Categorical data are presented as frequencies and proportions. We regarded other causes of death as competing events in our analysis of competing risks. When there is a competiting risk, the outcome is not only survival, death. Cumulative incidence function, CIFk(t) = Pr(T [?] t, D = k), represents the probability of the k event before time t and other types of events[6].The comparison between the cumulative incidences of the groups is checked by the Gray test[6]. Univariate analysis was performed using the cumulative incidence function (CIF) to show the probability of each event and Gray's test to estimate the difference in the CIF between groups[7]. Multivariate analysis with the Fine-Gray model was used to identify factors affecting the cumulative incidence of stage II. The Fine-Gray model is designed to fit the cumulative incidence of events of interest[8]. It is suitable for personal risk prediction research, tends to estimate disease risk and prognosis, and is suitable for establishing clinical prediction models and risk scores[9]. We also compared the results from a Cox regression model with those from the Fine-Gray model. The cumulative hazard was compared with a competing-risks model constructed using Kaplan-Meier estimation. All statistical analyses were performed using SPSS (version 24.0, SPSS), and R statistical software (version 3.5.0; https ://www.r-proje ct.org/). The "cmprsk" R package was used to construct the model. All statistical tests were two-sided, with $P < .05$ considered to be indicative of statistical significance.

## 3 Results

### 3.1 Patient characteristics

Of the 9389 eligible patients, 1318 died of other causes such as other cancers, suicide, and accidents, accounting for 14.0% of the total. Death due to other reasons was considered a competing event; 1144 died of stage II of colon cancer, accounting for 12.2% of the total. Of the 7363 white race patients, 881 died of stage II of colon cancer, accounting for 12.0% of the total. 1079 died of other causes, accounting for 14.7% of the total. Of the 4768 male patients, 562 died of stage II of colon cancer, accounting for 11.2% of the total. 735 died of other causes, accounting for 15.4% of the total. Of the 6175 proximal colon patients, 711 died of stage II of colon cancer, accounting for 11.5% of the total. 930 died of other causes, accounting for 15.1% of the total. Of the 7715 gradeI/II patients, 901 died of stage II of colon cancer, accounting for 11.6% of the total. 1057 died of other causes, accounting for 13.7% of the total. Of the 7987 stage T3 patients, 789 died of stage II of colon cancer, accounting for 9.9% of the total. 1159 died of other causes, accounting for 14.5% of the total. Of the number of lymph nodes (nLN<12) patients, 941 died of stage II of colon cancer, accounting for 11.1% of the total. 1152 died of other causes, accounting for 13.5% of the total. The results are provided in detail in Table 1.

### 3.2 Univariate analysis of the prognosis of stage II of colon cancer

The univariate analysis included Gray's test and the CIF. When competing risks were present, the results of Gray's test showed that race, tumor site, grade, T-stage, and the number of regional lymph nodes exerted statistically significant effects on stage II of colon cancer ($P < .05$). The CIF for almost all variables increased

over 1, 3, and 5 years, and was higher for black race patients, distal colon, gradeIII/IV, stage T4, and the number of regional lymph nodes exerted (nLN[?]12). The data are listed in detail in Table 2.

### 3.3 Multivariate analysis of the prognosis of stage II of colon cancer

When competing events were present, we included variables that were statistically significant in the univariate analysis in the Fine-Gray model. The multivariate analysis indicated that tumor site, grade, T-stage, and the number of regional lymph nodes exerted were significantly associated with survival. The data are listed in detail in Table 3, which includes the results from the multivariate Cox regression for comparison.

### 3.4 Comparative analysis

We compared the results from classical Kaplan-Meier curve analysis with the cumulative risk rate of the competing-risks model, which revealed that only Kaplan-Meier curve analysis led to an overestimation of the cumulative risk of the patient. The results show that, in fact, when there is a risk of competition, the cumulative risk of stage II of colon cancer patients is not as high as the cumulative risk of the K-M method. The cumulative incidence due to death from other causes for the same survival time was higher than that from stage II of colon cancer alone. If death from other causes is treated as censored, it will have a greater impact on the results. As can be seen from Figure 2.

## 4 Discussion

Single endpoints are rarely observed in medical research, with instead multiple endpoints that compete with each other commonly being present[10,11]. The occurrence of competing events hinders analyses of the occurrence of ending events of interest in a study. Previous studies have widely used Kaplan-Meier estimates of survival curves and Cox regression models to describe survival trends and identify important prognostic factors[10]. In the real world, the research object not only experiences one type of event, but different types of ending events affect each other, that is, form competing events. The statistical model for processing data with competitive events is called the "competing risk model." Survival data positive events usually include all-cause death and cause-specific death. When the study does not involve competing risks, K-M, COX regression method can be used for research. However, medical research generally has competing risks. When discussing specific causes of death, the traditional method may overestimate the cumulative incidence of each variable. It is therefore necessary to use the competing-risks model to deal with multiple end events[12,13]. In our study, competing risk analysis did not consider events due to stage II colon cancer death. It also considers events that die for other reasons and the effects of events.

In this study, We conducted an in-depth analysis of the data of patients with stage II colon cancer after tumor resection in the US SEER database. We found that tumor site, degree of tumor differentiation, T stage, and number of lymph node are independent factors affecting the prognosis of patients with stage II colon cancer. Our study included 9389 stage II colon cancer patients who had died between 2010 and 2015, with 1318 dying of other causes such as other cancers, suicides, and accidents, while 1144 had died of their stage II colon cancer. Our Fine-Gray regression analysis revealed that the Cox regression model overestimated the risk of white race (HR=1.151, 95% CI = 0.937-1.415), black race (HR=1.546, 95% CI=1.206-1.982), gradeIII/IV (HR=1.261,95%=1.092-1.456), T4 stage(HR=2.800,95%=2.468-3.176), the number of regional lymph nodes exerted(nLN<12) (HR=2.091,95%=1.793-2.438). These observations indicate that the relative risk of a patient dying from penile cancer when a competing event is present is different from when considering only a single endpoint event. This is the first study to use the competing-risks model to analyze the survival of patients with stage II colon cancer. When a competing event exists, the incidence of events of interest in the cumulative risk model is conditional on the composite event rate of all events of interest and those competing events, whereas the Kaplan-Meier estimation is only conditional on the incidence of events of interest. We compared the results from classical Kaplan-Meier curve analysis with the cumulative risk rate of the competing-risks model. When a competing event is treated as censored data, using the Kaplan-Meier method to calculate the cumulative risk results in a larger effect than the cumulative risk calculated using the competing-risks model, thereby overestimating the actual situation.

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 stage II colon cancer, and these factors may seriously affect the prognosis of patients with stage 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 stage II colon cancer.

In conclusion, this study established a competing-risks analysis model for the first time based on the SEER database for risk assessments of stage II colon cancer patients. The obtained results may help clinicians for clinical decision-making consultation and guidance for patients with stage II colon cancer.

**Reference**


1. Eguchi T, Bains S, Lee M, et al. Impact of increasing age on causespecific mortality and morbidity in patients with stage I non– small-cell lung cancer: a competing risks analysis. J Clin Oncol. 2017;35:281-290.
2. David HA. The theory of competing risks. Australian J Statistics. 1976;3:101-110.
3. van Walraven C, McAlister FA. Competing risk bias was common in Kaplan-Meier risk estimates published in prominent medical journals. J Clin Epidemiol. 2016;69:170-173.e8.
4. 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.
5. 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.
6. Haller B, Schmidt G, Ulm K. Applying competing risks regression models: an overview. Lifetime Data Anal. 2013;19:33-58.
7. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. Stat Med. 2007;26(11):2389-2430.
8. Fine JP, Gray R. A proportional hazards model for the subdistribution of a competing risk. J Am Statistical Assoc. 1999;94(446):496-509.
9. Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. Am J Epidemiol. 2009;170(2):244-256. 10. Scrucca L, Santucci A, Aversa F. Competing risk analysis using R: an easy guide for clinicians. Bone Marrow Transplant. 2007;40:381-387. 11. Moriña D, Navarro A. Competing risks simulation with the survsim R package. Commun Stat-Simul Computation. 2017;46:5712-5722. 12. Southern DA, Faris PD, Brant R, et al. Kaplan–Meier methods yielded misleading results in competing risk scenarios. J Clin Epidemiol. 2006;59:1110-1114. 13. Jepsen P, Vilstrup H, Andersen PK. The clinical course of cirrhosis: the importance of multistate models and competing risks analysis. Hepatology. 2015;62:292-302.


**Hosted file**

`table.docx` available at https://authorea.com/users/336139/articles/461909-competing-risks-model-for-predicting-the-prognosis-of-stage-%E2%85%B1-colon-cancer

**Hosted file**

`figure.docx` available at https://authorea.com/users/336139/articles/461909-competing-risks-model-for-predicting-the-prognosis-of-stage-%E2%85%B1-colon-cancer