Predicting Survival Outcome of Patients With Colorectal Cancer and Only Lung Metastasis: A Population-Based Real-World Study

Hui Tang

Peking Union Medical College Hospital <https://orcid.org/0000-0001-8870-6726>

Jian-Feng Zhou (✉ ZhouJF@pumch.cn)

Peking Union Medical College Hospital

Chun-Mei Bai

Peking Union Medical College Hospital

Research article

Keywords: colorectal cancer, pulmonary metastasis, prognostic evaluation, nomogram, SEER database

Posted Date: October 7th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-80102/v1>

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

[Read Full License](#)

Abstract

Background: Patients with metastatic colorectal cancer (mCRC) have a poor prognosis, but lung metastasis (LM) generally has a relatively desirable survival outcome. However, clinicians have had few tools for estimating the probability of survival in patients with colorectal cancer (CRC) and only LM (OLM). The present study aimed to develop nomograms estimating survival probability for patients with CRC and OLM.

Methods: Data from the Surveillance, Epidemiology, and End Results (SEER) database of patients with CRC between 2010 and 2014 were retrieved for retrospective analysis. Patients with OLM diagnosed between 2010 and 2014, except for 2012 ($n = 1,118$) were used to conduct multivariate Cox analysis to identify independent prognostic factors. Nomograms estimating 1- and 3-year overall survival (OS) and cancer-specific survival (CSS) were developed. The nomograms were internally validated for concordance index (C-index), calibration plots, receiver operating characteristic (ROC) curves, and were also externally validated with independent patients diagnosed in 2012 ($n = 261$).

Results: Age, marital status, tumor location, tumor size, T and N stage, CEA, tumor deposit, histological grade, primary or metastatic tumor surgery, chemotherapy, radiotherapy, and income were found to be independently associated with OS and/or CSS. The nomograms were constructed based on these prognostic factors. The C-index were high in internal validation (0.736 for OS and 0.741 for CSS) and external validation (0.656 for OS and 0.663 for CSS). Internal and external calibration plots and ROC curves demonstrated a good agreement between actual observation and nomogram prediction.

Conclusions: The nomograms individually predict OS and CSS of patients with CRC and OLM and could aid in the personalized prognostic evaluation and clinical decision-making.

1. Introduction

Colorectal cancer (CRC) is the second leading cause of cancer death and the third most commonly diagnosed cancer worldwide, the incidence rates tend to rise consistently with social development [1]. With multidisciplinary treatment, including surgery for primary and metastasis tumor, systemic therapy and radiotherapy (mainly for rectal cancer), the 5-year relative survival rate is 65% for patients with CRC, but declines to 12% for metastatic CRC (mCRC) [2].

In terms of clinicopathological factors, old age, advanced tumor-node-metastasis (TNM) stage, and high histological grade are generally considered to be poor prognostic factors for CRC. Moreover, primary site, tumor size, tumor deposit, carcinoembryonic antigen (CEA), treatment, and some demographic factors can also influence the prognosis of CRC [3]. Patients with mCRC have a poor prognosis, but those with only lung metastasis (OLM) generally have a relatively desirable prognosis [4]. Why patients with OLM have a better prognosis and how to predict accurately the prognosis of patients with OLM are of interest to clinicians. However, clinicians have had few tools for estimating the probability of survival in patients with CRC and OLM.

Nomogram, a statistical prediction tool that can incorporate and quantify the selected prognostic factors to estimate the survival probability for an individual patient, has been widely demonstrated in cancers including CRC, lung cancer, and gastric cancer [5–7]. However, to our knowledge, the nomograms predicting the overall survival (OS) and cancer-specific survival (CSS) in patients with CRC and OLM have not been reported. Established in 1973, the Surveillance, Epidemiology, and End Results (SEER) program provides a wide range of clinicopathological and follow-up information of cancer patients, covering about 28% of the population in the United States [8]. Using the SEER database, we constructed nomograms estimating survival probability for patients with CRC and OLM. Furthermore, the internal and external validation for nomograms were performed.

2. Methods

2.1. Data source and selection

Patient data were retrieved from the SEER database using the SEER*stat software, version 8.3.6 (<https://seer.cancer.gov/seerstat/>). The flow chart of the case selection is shown in Fig. 1. The inclusion criteria were as follows: (1) stage IV colorectal cancer patients; (2) positive histological confirmation; (3) because detailed information about distant metastatic sites was not available before 2010 and the follow-up information was updated to November 2018, the present study was restricted to patients who were diagnosed between 2010 and 2014 to ensure a minimal follow-up length of three years. The exclusion criteria were as follows: (1) patients who had multiple tumors in their lifetime; (2) unknown important and easily accessible information in clinical practice including age at diagnosis, race, and marital status; (3) unknown survival time; (4) unknown or no metastasis in specific metastatic sites (bone, brain, liver, and lung). All the included patients were used to analyze the risk factors for lung metastasis. After that, patients with only bone, brain, liver, or lung metastasis were divided to explore the effect of specific metastatic sites on OS. Furthermore, patients with OLM were divided into training or validation cohorts based on year of diagnosis for further analysis.

2.2. Study variables

Clinical variables of CRC patients were extracted, including age, race, sex, marital status, TNM stage, tumor location, tumor size, specific metastatic sites (bone, brain, liver, and lung), CEA, tumor deposit, perineural invasion, histological grade, treatment, vital status, survival time, corresponding death causes, the status of education and income in the county patients came from. In our study, except for the white and the black, other race was recorded as “Other”. “Married (including common law)” was recorded as “Married”, other marital status was recorded as “Single/unmarried”. Tumor location was defined either as “Right colorectal”, “Left colorectal”, or “Unknown”. “Right colorectal” included International Classification of Diseases 10th Revision (ICD-10) codes C180 to 184, and “Left colorectal” included ICD-10 codes C185 to 187, C199 and C209. “Unknown” included ICD-10 code C188, C189. Considering it is difficult to distinguish T0-3 stages in patients without primary tumor resection, T0-3 stages were integrated into one. The status of education and income defined either as “Low” or “High”, meaning that patients came from

counties with lower/higher than median education or income level. Considering the no precise survival days were available, survival time of 0 months was recorded as 0.5 months to patients who died within one month of diagnosis but who did not reach the one-month threshold [9]. OS was defined as the time period from the diagnosis to the death caused by any cause or last follow-up, while CSS was defined as the time period from the diagnosis to the death caused by CRC.

2.3. Statistical analysis

The chi-square tests were used to compare the clinicopathological characteristics among training cohort and validation cohort or with and without lung metastasis. Multivariate binary logistic regression analyses were performed to identify risk factors of lung metastasis. Differences in OS and CSS were compared between patients with specific metastatic sites. Kaplan-Meier (KM) analyses and log-rank tests were used to assess this effect. Univariate and multivariate Cox proportional hazards regression models were performed to analyze the hazard ratios (HR) with corresponding 95% confidence intervals (CI) of variables associated with OS and CSS. Based on multivariate Cox analyses in the training cohort, nomograms were developed and evaluated by the concordance index (C-index), receiver operating characteristics (ROC), and calibration curves. The C-index was used to evaluate the predictive discrimination ability and accuracy of the overall nomogram: the higher the C-index, the better model's discrimination ability. The ROC curves are similar to the C-index, but are considered less suitable for use with censored data. The calibration curves were used to evaluate the accuracy of nomogram-predicted survival probability comparing with the actual survival probability. Closer distances from the points to the line indicate higher prediction accuracy of calibration curves. To verify the universal applicability of the model, nomograms were both internally and externally validated. All statistical analyses were performed in R version 3.6.1 (<http://www.r-project.org/>) or SPSS software version 21.0 (IBM Corporation). For all of the analyses, a two-tailed value of $p < 0.05$ was defined as significant.

3. Results

3.1. Patients characteristics

As shown in Fig. 1, a total of 992,325 patients with diagnoses of CRC were registered in the SEER database from 1969 to 2017. Among the 19,161 patients met the selection criteria, 14,360 patients were diagnosed CRC with only bone, brain, liver, or lung metastasis. Furthermore, patients with OLM ($n = 1,379$) were divided into two cohorts based on year of diagnosis for further analysis: patients diagnosed in 2012 were divided into validation cohort ($n = 261$), other patients were divided into training cohort ($n = 1,118$). The demographic and clinicopathologic characteristics of patients with OLM were listed in Table 1. Chi-square tests showed that there was no significant difference in patients' demographic and clinicopathologic characteristics between training and validation cohorts.

Table 1

Baseline characteristics of the included patients with colorectal cancer and only lung metastasis.

Characteristic	Total n = 1,379 No. (%)	Training cohort ^a n = 1,118 No. (%)	Validation cohort ^b n = 261 No. (%)	Pearson χ^2 Pvalue
Age, years old				0.439
0–49	196 (14.21)	167 (14.94)	29 (11.11)	
50–64	472 (34.23)	377 (33.72)	95 (36.40)	
65–79	482 (34.95)	390 (34.88)	92 (35.25)	
≥80	229 (16.61)	184 (16.46)	45 (17.24)	
Sex				0.823
Male	706 (51.20)	574 (51.34)	132 (50.57)	
Female	673 (48.80)	544 (48.66)	129 (49.43)	
Race				0.286
White	1,020 (73.97)	825 (73.79)	195 (74.71)	
Black	214 (15.52)	169 (15.12)	45 (17.24)	
Other ^c	145 (10.51)	124 (11.09)	21 (8.05)	
Marital status				0.223
Married	665 (48.22)	548 (49.02)	117 (44.83)	
Single/unmarried ^d	714 (51.78)	570 (50.98)	144 (55.17)	
Location				0.412
Right colorectal	360 (26.11)	290 (25.94)	70 (26.82)	
Left colorectal	936 (67.88)	765 (68.43)	171 (65.52)	
Unknown	83 (6.02)	63 (5.64)	20 (7.66)	
Tumor size				0.914
<3 cm	106 (7.67)	84 (7.51)	22 (8.43)	
≥3 & <5 cm	318 (23.06)	259 (23.17)	59 (22.61)	
≥5 cm	515 (37.35)	421 (37.66)	94 (36.02)	

Characteristic	Total n = 1,379 No. (%)	Training cohort a n = 1,118 No. (%)	Validation cohort b n = 261 No. (%)	Pearson χ^2 Pvalue
Unknown	440 (31.91)	354 (31.66)	86 (32.95)	
T stage				0.069
T0-3	727 (52.72)	606 (54.20)	121 (46.36)	
T4	355 (25.74)	277 (24.78)	78 (29.89)	
Tx	297 (21.54)	235 (21.02)	62 (23.75)	
N stage				0.633
N0	490 (35.53)	398 (35.60)	92 (35.25)	
N1	490 (35.53)	402 (35.96)	88 (33.72)	
N2	259 (18.78)	210 (18.78)	49 (18.77)	
Nx	140 (10.15)	108 (9.66)	32 (12.26)	
CEA				0.705
Positive	677 (49.09)	548 (49.02)	129 (49.43)	
Negative	256 (18.56)	212 (18.96)	44 (16.86)	
Unknown	446 (32.34)	358 (32.02)	88 (33.72)	
Tumor deposit				0.662
Positive	202 (14.65)	166 (14.85)	36 (13.79)	
Negative	451 (32.70)	370 (33.09)	81 (31.03)	
Unknown ^e	726 (52.65)	582 (52.06)	144 (55.17)	
Perineural invasion				0.275
Positive	173 (12.55)	134 (11.99)	39 (14.94)	
Negative	600 (43.51)	496 (44.36)	104 (39.85)	
Unknown ^e	606 (43.94)	488 (43.65)	118 (45.21)	
Grade				0.605
I	74 (5.37)	59 (5.28)	15 (5.75)	
II	760 (55.11)	624 (55.81)	136 (52.11)	

Characteristic	Total n = 1,379 No. (%)	Training cohort ^a n = 1,118 No. (%)	Validation cohort ^b n = 261 No. (%)	Pearson χ^2 Pvalue
III	208 (15.08)	171 (15.30)	37 (14.18)	
IV	38 (2.76)	31 (2.77)	7 (2.68)	
Unknown	299 (21.68)	233 (20.84)	66 (25.29)	
Primary tumor surgery				0.823
Yes	706 (51.20)	574 (51.34)	132 (50.57)	
No/unknown ^f	673 (48.80)	544 (48.66)	129 (49.43)	
Metastatic tumor surgery				0.795
Yes	165 (11.97)	135 (12.08)	30 (11.49)	
No/unknown ^f	1,214 (88.03)	983 (87.92)	231 (88.51)	
Chemotherapy				0.336
Yes	938 (68.02)	767 (68.61)	171 (65.52)	
No/unknown ^g	441 (31.98)	351 (31.40)	90 (34.48)	
Radiotherapy				0.749
Yes	312 (22.63)	251 (22.45)	61 (23.37)	
No/unknown ^g	1,067 (77.37)	867 (77.55)	200 (76.63)	
Education ^h				0.767
Low	497 (36.04)	405 (36.23)	92 (35.25)	
High	882 (63.96)	713 (63.77)	169 (64.75)	
Income ^h				0.905
Low	545 (39.52)	441 (39.45)	104 (39.85)	
High	834 (60.48)	677 (60.55)	157 (60.15)	
Abbreviations: CEA, carcinoembryonic antigen.				
^a Patients diagnosed between 2010 and 2014, except for 2012.				

Characteristic	Total n = 1,379 No. (%)	Training cohort ^a n = 1,118 No. (%)	Validation cohort ^b n = 261 No. (%)	Pearson χ^2 Pvalue
^b Patients diagnosed in 2012.				
^c American Indian/Alaska native and Asian/Pacific Islander.				
^d Single (never married), unmarried or domestic partner, widowed divorced and separated were included.				
^e Including patients who had no surgical resection of the primary site.				
^f Since there were fewer patients with unknown surgical status, the two were integrated.				
^g Can't tell the difference between "No" and "Unknown".				
^h Patients came from counties with lower/higher than median education or income level.				

3.2. Factors associated with lung metastasis

As shown in Table 2, chi-square tests indicated that lung metastasis (LM) was significantly associated with age, race, marital status, tumor location, tumor size, T and N stage, other organs metastases (bone, brain, and liver), serum CEA level, tumor deposit, perineural invasion, and histological grade. Significant factors identified by chi-square tests were further explored in multivariate analysis, which showed that old age, non-white descent, unmarried status, left colorectal tumor, large size of the tumor, advanced T and N stage, liver metastasis, positive serum CEA, positive tumor deposit and high histological grade were independent risk factors for LM.

Table 2

Multivariate binary logistic regression analysis of stage IV colorectal cancer patients' characteristics classified by lung metastasis at diagnosis (n = 19,161).

Variable	Lung metastasis		Pearson χ^2 Pvalue	Multivariate analysis		
	Yes; n = 5,601	No; n = 13,560		OR	95% CI	Pvalue
	No. (%)	No. (%)				
Age, years old			< 0.001			
0–49	819 (26.45)	2,277 (73.55)		1.000		
50–64	2,155 (28.77)	5,335 (71.23)		1.148	(1.034,1.275)	0.010
65–79	1,851 (30.95)	4,130 (69.05)		1.272	(1.141,1.418)	< 0.001
≥80	776 (29.92)	1,818 (70.08)		1.148	(1.005,1.310)	0.042
Sex			0.156	NI		
Male	3,084 (28.82)	7,618 (71.18)				
Female	2,517 (29.76)	5,942 (70.24)				
Race			< 0.001			
White	4,077 (28.41)	10,275 (71.59)		1.000		
Black	940 (30.92)	2,100 (69.08)		1.137	(1.033,1.251)	0.008
Other ^a	584 (33.01)	1,185 (66.99)		1.230	(1.094,1.383)	0.001

Abbreviations: CEA, carcinoembryonic antigen; CI, confidence interval; NI, not included; OR, odds ratio.

^a American Indian/Alaska native and Asian/Pacific Islander.

^b Single (never married), unmarried or domestic partner, widowed divorced and separated were included.

^c Including patients who had no surgical resection of the primary site.

^d Can't tell the difference between "No" and "Unknown".

^e Patients came from counties with lower/higher than median education or income level.

Variable	Lung metastasis		Pearson χ^2 Pvalue	Multivariate analysis		
	Yes; n = 5,601	No; n = 13,560		OR	95% CI	Pvalue
	No. (%)	No. (%)				
Marital status			< 0.001			
Married	2,766 (27.51)	7,288 (72.49)		1.000		
Single/unmarried ^b	2,835 (31.13)	6,272 (68.87)		1.104	(1.029,1.185)	0.006
Location			< 0.001			
Right colorectal	1,621 (23.88)	5,167 (76.12)		1.000		
Left colorectal	3,548 (31.92)	7,568 (68.08)		1.297	(1.199,1.403)	< 0.001
Unknown	432 (34.37)	8,25 (65.63)		1.109	(0.953,1.290)	0.181
Tumor size			< 0.001			
<3 cm	311 (24.74)	946 (75.26)		1.000		
≥3 & <5 cm	1,061 (25.49)	3,102 (74.51)		1.285	(1.087,1.519)	0.003
≥5 cm	1,969 (26.55)	5,446 (73.45)		1.338	(1.141,1.568)	< 0.001
Unknown	2,260 (35.73)	4,066 (64.27)		1.384	(1.175,1.629)	< 0.001
T stage			< 0.001			

Abbreviations: CEA, carcinoembryonic antigen; CI, confidence interval; NI, not included; OR, odds ratio.

^a American Indian/Alaska native and Asian/Pacific Islander.

^b Single (never married), unmarried or domestic partner, widowed divorced and separated were included.

^c Including patients who had no surgical resection of the primary site.

^d Can't tell the difference between "No" and "Unknown".

^e Patients came from counties with lower/higher than median education or income level.

Variable	Lung metastasis		Pearson χ^2 Pvalue	Multivariate analysis		
	Yes; n = 5,601	No; n = 13,560		OR	95% CI	Pvalue
	No. (%)	No. (%)				
T0-3	2,465 (26.57)	6,814 (73.43)		1.000		
T4	1,251 (26.40)	3,488 (73.60)		1.087	(0.993,1.191)	0.072
Tx	1,885 (36.65)	3,258 (63.35)		1.198	(1.087,1.320)	< 0.001
N stage			< 0.001			
N0	1,894 (30.93)	4,229 (69.07)		1.000		
N1	1,906 (30.00)	4,447 (70.00)		1.240	(1.135,1.356)	< 0.001
N2	940 (21.52)	3,429 (78.48)		1.108	(0.984,1.248)	0.090
Nx	861 (37.18)	1,455 (62.82)		1.138	(1.015,1.276)	0.027
Bone metastasis			< 0.001			
Yes	582 (43.53)	755 (56.47)		1.000		
No	5,019 (28.16)	12,805 (71.84)		0.920	(0.804,1.053)	0.226
Brain metastasis			< 0.001			
Yes	156 (48.30)	167 (51.17)		1.000		

Abbreviations: CEA, carcinoembryonic antigen; CI, confidence interval; NI, not included; OR, odds ratio.

^a American Indian/Alaska native and Asian/Pacific Islander.

^b Single (never married), unmarried or domestic partner, widowed divorced and separated were included.

^c Including patients who had no surgical resection of the primary site.

^d Can't tell the difference between "No" and "Unknown".

^e Patients came from counties with lower/higher than median education or income level.

Variable	Lung metastasis		Pearson χ^2 Pvalue	Multivariate analysis		
	Yes; n = 5,601	No; n = 13,560		OR	95% CI	Pvalue
	No. (%)	No. (%)				
No	5,445 (28.90)	13,393 (71.10)		1.318	(0.999,1.739)	0.051
Liver metastasis			< 0.001			
Yes	4,061 (23.48)	13,233 (76.52)		1.000		
No	1,540 (82.49)	327 (17.51)		18.091	(15.833,20.670)	< 0.001
CEA			< 0.001			
Positive	3,459 (30.42)	7,912 (69.58)		1.000		
Negative	488 (24.53)	1,501 (75.47)		0.606	(0.531,0.692)	< 0.001
Unknown	1,654 (28.51)	4,147 (71.49)		0.838	(0.773,0.904)	< 0.001
Tumor deposit			< 0.001			
Positive	673 (21.74)	2,422 (78.26)		1.000		
Negative	1,309 (21.61)	4,747 (78.39)		0.929	(0.823,1.048)	0.232
Unknown ^c	3,619 (36.15)	6,391 (63.85)		1.726	(1.506,1.979)	< 0.001
Perineural invasion			< 0.001			

Abbreviations: CEA, carcinoembryonic antigen; CI, confidence interval; NI, not included; OR, odds ratio.

^a American Indian/Alaska native and Asian/Pacific Islander.

^b Single (never married), unmarried or domestic partner, widowed divorced and separated were included.

^c Including patients who had no surgical resection of the primary site.

^d Can't tell the difference between "No" and "Unknown".

^e Patients came from counties with lower/higher than median education or income level.

Variable	Lung metastasis		Pearson χ^2 Pvalue	Multivariate analysis		
	Yes; n = 5,601	No; n = 13,560		OR	95% CI	Pvalue
	No. (%)	No. (%)				
Positive	560 (21.88)	1,999 (78.12)		1.000		
Negative	1,977 (25.43)	5,796 (74.57)		0.994	(0.877,1.126)	0.920
Unknown ^c	3,064 (34.70)	5,765 (65.30)		1.073	(0.930,1.237)	0.334
Grade			< 0.001			
I	242 (30.63)	548 (69.37)		1.000		
II	2,854 (28.58)	7,133 (71.42)		1.028	(0.863,1.225)	0.756
III	803 (23.78)	2,574 (76.22)		0.757	(0.626,0.917)	0.004
IV	130 (21.14)	485 (78.86)		0.790	(0.601,1.040)	0.094
Unknown	1,572 (35.79)	2,820 (64.21)		1.051	(0.874,1.264)	0.597
Education ^e			0.616	NI		
Low	2,033 (29.45)	4,870 (70.55)				
High	3,568 (29.11)	8,690 (70.89)				
Income ^e			0.404	NI		

Abbreviations: CEA, carcinoembryonic antigen; CI, confidence interval; NI, not included; OR, odds ratio.

^a American Indian/Alaska native and Asian/Pacific Islander.

^b Single (never married), unmarried or domestic partner, widowed divorced and separated were included.

^c Including patients who had no surgical resection of the primary site.

^d Can't tell the difference between "No" and "Unknown".

^e Patients came from counties with lower/higher than median education or income level.

Variable	Lung metastasis		Pearson χ^2 Pvalue	Multivariate analysis		
	Yes; n = 5,601	No; n = 13,560		OR	95% CI	Pvalue
	No. (%)	No. (%)				
Low	2,143 (29.58)	5,101 (70.42)				
High	3,458 (29.02)	8,459 (70.98)				
Abbreviations: CEA, carcinoembryonic antigen; CI, confidence interval; NI, not included; OR, odds ratio.						
^a American Indian/Alaska native and Asian/Pacific Islander.						
^b Single (never married), unmarried or domestic partner, widowed divorced and separated were included.						
^c Including patients who had no surgical resection of the primary site.						
^d Can't tell the difference between "No" and "Unknown".						
^e Patients came from counties with lower/higher than median education or income level.						

3.3. Differences in OS and CSS between patients with specific metastatic sites

To confirm that different metastatic sites would affect the survival outcome of CRC patients, we further grouped the patients according to the site of metastasis. In this process, we only included the patients with CRC and only bone (n = 225), brain (n = 91), liver (n = 12,665) or lung metastasis (n = 1,379). According to the Kaplan-Meier curves and log-rank analyses of OS and CSS (Figs. 2), among patients with CRC and only bone, brain, liver, or lung metastasis, those with OLM had the best prognosis. Conversely, CRC patients with bone or brain metastasis had a poorer prognosis, regardless of OS or CSS, this is consistent with previous studies [4].

3.4. Establishment of the nomogram

Among patients with CRC and OLM, age, marital status, tumor location, tumor size, T and N stage, serum CEA level, tumor deposit, perineural invasion, histological grade, primary and metastatic tumor surgery, chemotherapy, radiotherapy, and income were significantly associated with OS by univariate analysis in the training cohort (Table 3). However, further multivariate analysis indicated that only age, marital status, tumor location, T and N stage, serum CEA level, tumor deposit, histological grade, primary tumor surgery, chemotherapy, radiotherapy, and income were significantly associated with OS. Based on the multivariate results, a nomogram of 1- and 3-year OS was established with the independent variables (Fig. 3A). With the same procedure, a nomogram of CSS was established (Table 4 and Fig. 3B).

Table 3
Univariate and multivariate analyses of overall survival in the training cohort (n = 1,118).

Characteristic	Univariate analysis P-value	Multivariate analysis		
		HR	95% CI	Pvalue
Age, years old	< 0.001			
0–49		1.000		
50–64		0.967	(0.771,1.213)	0.770
65–79		1.079	(0.857,1.359)	0.517
≥80		1.389	(1.066,1.810)	0.015
Sex	0.278	NI		
Male				
Female				
Race	0.275	NI		
White				
Black				
Other ^a				
Marital status	< 0.001			
Married		1.000		
Single/unmarried ^b		1.342	(1.162,1.549)	< 0.001
Location	< 0.001			
Right colorectal		1.000		
Left colorectal		0.813	(0.682,0.968)	0.020
Unknown		1.251	(0.900,1.738)	0.182
Tumor size	< 0.001			
<3 cm		1.000		
≥3 & <5 cm		1.170	(0.852,1.607)	0.331
≥5 cm		1.141	(0.839,1.551)	0.399
Unknown		1.227	(0.900,1.674)	0.196
T stage	< 0.001			

Characteristic	Univariate analysis P-value	Multivariate analysis		
		HR	95% CI	Pvalue
T0-3		1.000		
T4		1.453	(1.219,1.732)	< 0.001
Tx		1.226	(0.991,1.516)	0.060
N stage	< 0.001			
N0		1.000		
N1		1.032	(0.866,1.231)	0.722
N2		1.364	(1.067,1.744)	0.013
Nx		0.852	(0.655,1.108)	0.232
CEA	< 0.001			
Positive		1.000		
Negative		0.698	(0.573,0.850)	< 0.001
Unknown		0.735	(0.624,0.865)	< 0.001
Tumor deposit	< 0.001			
Positive		1.000		
Negative		0.741	(0.588,0.934)	0.011
Unknown ^c		0.759	(0.545,1.058)	0.104
Perineural invasion	< 0.001			
Positive		1.000		
Negative		0.953	(0.743,1.223)	0.708
Unknown ^c		1.114	(0.832,1.491)	0.469
Grade	< 0.001			
I		1.000		
II		1.105	(0.795,1.538)	0.552
III		1.705	(1.187,2.450)	0.004
IV		1.647	(0.989,2.745)	0.055
Unknown		1.041	(0.728,1.490)	0.824
Primary tumor surgery	< 0.001			

Characteristic	Univariate analysis P-value	Multivariate analysis		
		HR	95% CI	Pvalue
Yes		1.000		
No/unknown ^d		2.265	(1.701,3.018)	< 0.001
Metastatic tumor surgery	< 0.001			
Yes		1.000		
No/Unknown ^d		1.236	(0.967,1.578)	0.090
Chemotherapy	< 0.001			
Yes		1.000		
No/unknown ^e		2.459	(2.085,2.901)	< 0.001
Radiotherapy	< 0.001			
Yes		1.000		
No/unknown ^e		1.278	(1.062,1.538)	0.009
Education ^f	0.354	NI		
Low				
High				
Income ^f	0.031			
Low		1.000		
High		0.814	(0.705,0.939)	0.005
Abbreviations: CEA, carcinoembryonic antigen; CI, confidence interval; HR hazard ratio; NI, not included.				
^a American Indian/Alaska native and Asian/Pacific Islander.				
^b Single (never married), unmarried or domestic partner, widowed divorced and separated were included.				
^c Including patients who had no surgical resection of the primary site.				
^d Since there were fewer patients with unknown surgical status, the two were integrated.				
^e Can't tell the difference between "No" and "Unknown".				
^f Patients came from counties with lower/higher than median education or income level.				

Table 4

Univariate and multivariate analyses of cancer-specific survival in the training cohort (n = 1,118).

Characteristic	Univariate analysis P-value	Multivariate analysis		
		HR	95% CI	Pvalue
Age, years old	< 0.001			
0–49		1.000		
50–64		0.910	(0.720,1.149)	0.428
65–79		1.004	(0.792,1.274)	0.973
≥80		1.345	(1.024,1.768)	0.033
Sex	0.240	NI		
Male				
Female				
Race	0.304	NI		
White				
Black				
Other ^a				
Marital status	< 0.001			
Married		1.000		
Single/unmarried ^b		1.343	(1.156,1.561)	< 0.001
Location	< 0.001			
Right colorectal		1.000		
Left colorectal		0.784	(0.654,0.941)	0.009
Unknown		1.244	(0.882,1.756)	0.213
Tumor size	< 0.001			
<3 cm		1.000		
≥3 & <5 cm		1.158	(0.828,1.618)	0.392
≥5 cm		1.132	(0.819,1.564)	0.453
Unknown		1.245	(0.897,1.730)	0.191
T stage	< 0.001			

Characteristic	Univariate analysis P-value	Multivariate analysis		
		HR	95% CI	Pvalue
T0-3		1.000		
T4		1.475	(1.229,1.770)	< 0.001
Tx		1.213	(0.970,1.518)	0.091
N stage	< 0.001			
N0		1.000		
N1		1.039	(0.865,1.249)	0.681
N2		1.375	(1.065,1.775)	0.015
Nx		0.838	(0.636,1.105)	0.211
CEA	< 0.001			
Positive		1.000		
Negative		0.692	(0.563,0.851)	< 0.001
Unknown		0.754	(0.636,0.895)	0.001
Tumor deposit	< 0.001			
Positive		1.000		
Negative		0.717	(0.564,0.911)	0.006
Unknown ^c		0.726	(0.514,1.027)	0.070
Perineural invasion	< 0.001			
Positive		1.000		
Negative		0.923	(0.714,1.194)	0.542
Unknown ^c		1.088	(0.804,1.472)	0.584
Grade	< 0.001			
I		1.000		
II		1.073	(0.764,1.508)	0.684
III		1.751	(1.207,2.541)	0.003
IV		1.577	(0.929,2.677)	0.092
Unknown		1.018	(0.703,1.472)	0.926
Primary tumor surgery	< 0.001			

Characteristic	Univariate analysis P-value	Multivariate analysis		
		HR	95% CI	Pvalue
Yes		1.000		
No/unknown ^d		2.325	(1.718,3.146)	< 0.001
Metastatic tumor surgery	< 0.001			
Yes		1.000		
No/Unknown ^d		1.303	(1.004,1.689)	0.046
Chemotherapy	< 0.001			
Yes		1.000		
No/unknown ^e		2.371	(1.991,2.823)	< 0.001
Radiotherapy	< 0.001			
Yes		1.000		
No/unknown ^e		1.284	(1.056,1.560)	0.012
Education ^f	0.482	NI		
Low				
High				
Income ^f	0.081	NI		
Low				
High				
Abbreviations: CEA, carcinoembryonic antigen; CI, confidence interval; HR hazard ratio; NI, not included.				
^a American Indian/Alaska native and Asian/Pacific Islander.				
^b Single (never married), unmarried or domestic partner, widowed divorced and separated were included.				
^c Including patients who had no surgical resection of the primary site.				
^d Since there were fewer patients with unknown surgical status, the two were integrated.				
^e Can't tell the difference between “No” and “Unknown”.				
^f Patients came from counties with lower/higher than median education or income level.				

The point scale at the top of each nomogram was used first to give every prognostic factor a score, then adding up the scores of all variables. Finally, based on the total score, the point scale at the bottom of each nomogram was used to predict the survival probability of one individual patient.

3.5. Nomogram validation

The OS and CSS nomograms were validated both by internal and external cohorts. In the internal validation, the C-index, an indicator of a model's discrimination ability, was 0.736 (95%CI: 0.719–0.753) for the OS nomogram, while 0.741 (95%CI: 0.723–0.759) for the CSS nomogram. Besides, the area under ROC curve (AUC) for the 1- and 3-year OS nomogram were 0.811 and 0.799, respectively (Fig. 4A and 4B); while for the 1- and 3-year CSS nomogram were respectively 0.821 and 0.800 (Fig. 4C and 4D), suggesting high discrimination ability of the model. In the external validation, the C-index for the OS nomogram was 0.656 (95%CI: 0.615–0.697), while for the CSS nomogram was 0.663 (95%CI: 0.621–0.705). Besides, the AUC for the 1- and 3-year OS nomogram were 0.690 and 0.731, respectively (Fig. 4E and 4F); while for the 1- and 3-year CSS nomogram were 0.704 and 0.721, respectively (Fig. 4G and 4H). Furthermore, internal and external calibration plots both demonstrated that the nomograms well predicted 1- and 3-year OS and CSS (Fig. 5A-H), showing the excellent agreement between actual observation and nomogram prediction.

4. Discussion

Definitely, effective and precise prognostic prediction in patients with malignancy is important for clinical decision-making and scientific research. Previous studies had reported that factors such as age, sex, CEA, tumor size, T and N stage, metastatic sites, and treatment were significant predictors related to survival outcome of mCRC patients [10]. Intuitive and convenient prognostic tools are what clinicians need. However, the inclusion of more information and a more individualized prognosis model will certainly predict the prognosis of patients more accurately.

Using the SEER database which covers about 28% of the population in the United States, we were able to collect sufficient samples and therefore developed and validated nomograms that predicting 1- and 3-year OS and CSS of patients with CRC and OLM. In the current study, we identified age, marital status, tumor location, T and N stage, serum CEA level, tumor deposit, histological grade, primary tumor surgery, chemotherapy, radiotherapy, and income were significantly associated with OS. As for CSS, metastatic tumor surgery was a significant prognostic factor, but not income. The nomograms were created based on these significant prognostic factors. As for high-risk OLM patients, more aggressive treatments are recommended, and the follow-up interval is encouraged to be shortened appropriately to detect the occurrence of an endpoint event in the early stage.

Unsurprisingly, our results suggested that chemotherapy and surgical treatment were the main prognostic factors in CRC patients with OLM, and the primary site was also significantly associated with the prognosis. Indeed, it has been widely recognized that chemotherapy can improve the prognosis of mCRC patients, left-sided tumor location is also a predictor for better prognosis [11]. Moreover, surgery is still the

superior option for CRC patients who are able to undergo surgery, even for patients with mCRC. A comprehensive review including 26 clinical studies reported that the survival outcome of CRC patients with limited metastatic diseases were significantly improved by R0 resection of metastases [12]. Likewise, several retrospective studies and meta-analyses had demonstrated an association between primary tumor resection and improved OS in CRC patients with metastatic disease [13–16]. However, the data have not been entirely consistent across the published studies [17–19]. Of note, the correlation between radiotherapy and prognosis should be based on rectal cancer, and for the sake of simplicity, no distinction was made between left colon cancer and rectal cancer in the present study. Consistent with previous findings, our results suggested that sex, race, and tumor size had no effect on long-term survival for patients with LM from CRC [20]. Unlike it was clear that age was a prognostic factor in mCRC patients [21], whether sex, race, and tumor size had effects on the prognoses of patients with mCRC is still controversial. Some studies reported that white descent, male sex, larger primary tumor size predicts poorer survival [10, 22, 23].

In addition to prognostic nomograms, the present study identified several risk factors associated with LM from CRC. Combined with nomogram for prediction of the risk of LM from CRC [10, 20], clinicians and researchers can effectively identify patients at high risk of LM and monitor follow-up computed tomography closely. Moreover, we confirmed that different metastatic sites would affect the prognosis of CRC patients. Further exploration for why CRC patients with OLM had a better prognosis is worth looking forward to.

There were some limitations in our study. First, the nomograms were developed on the basis of retrospective data, prospective validation is needed. Second, some important information such as chemotherapy regimens, targeted therapy, molecular pathology, and genetic test were not recorded in the database. Third, the data of the timeline of other information such as CEA levels in patients were not available too. Finally, because of the limitations of our data, we didn't explore the effect of the factors, such as histological type and the number of positive lymph nodes, on the prognosis of mCRC patients. Further prospective studies with more critical information are warranted for model improvement and independent validation.

5. Conclusion

In summary, we developed nomograms predicting OS and CSS of patients with CRC and OLM individually. The nomograms showed good accuracy and applicability. Therefore, we recommend the simple and clear nomograms to be used as a convenient and effective tool for clinicians to evaluate the prognosis of individualized CRC patients with OLM and determine the treatment strategy.

Declarations

Funding

This work was supported by grants from CAMS Innovation Fund for Medical Sciences (No. 2016-I2M-1-001).

Competing interests

No potential conflict of interest was reported by the authors.

Acknowledgment

The authors would like to thank the Surveillance, Epidemiology, and End Results (SEER) database for the support.

Contributions

H T, JF Z, and CM B conceived of and designed the study. H T and JF Z performed literature search. H T generated the figures and Tables. H T and JF Z analyzed the data. H T wrote the manuscript, JF Z and CM B critically reviewed the manuscript. CM B supervised the research. All authors have read and approved the manuscript.

Ethics approval and consent to participate

We received permission to access the research data file in the SEER program from the National Cancer Institute, US (reference number 12317-Nov2019). Approval was waived by the local ethics committee, as SEER data is publicly available and de-identified.

Consent for publication

Not applicable.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA: a cancer journal for clinicians, 68 (2018) 394–424.
2. Miller KD, Nogueira L, Mariotto AB, Rowland JH, Yabroff KR, Alfano CM, Jemal A, Kramer JL, Siegel RL. Cancer treatment and survivorship statistics, 2019, CA: a cancer journal for clinicians, 69 (2019) 363–385.
3. Mahar AL, Compton C, Halabi S, Hess KR, Weiser MR, Groome PA. Personalizing prognosis in colorectal cancer: A systematic review of the quality and nature of clinical prognostic tools for survival outcomes. Journal of surgical oncology. 2017;116:969–82.
4. Prasanna T, Karapetis CS, Roder D, Tie J, Padbury R, Price T, Wong R, Shapiro J, Nott L, Lee M, Chua YJ, Craft P, Piantadosi C, Sorich M, Gibbs P, Yip D. The survival outcome of patients with metastatic

- colorectal cancer based on the site of metastases and the impact of molecular markers and site of primary cancer on metastatic pattern, *Acta oncologica* (Stockholm, Sweden), 57 (2018) 1438–1444.
5. Fernández Montes A, López López C, Argilés Martínez G, Páez López D, López Muñoz AM, García Paredes B, Gutiérrez Abad D, Castañón López C, Jiménez Fonseca P, Gallego Plazas J, López Doldán MC, Martínez de Castro E, Sánchez Cánovas M, Tobeña M, Puyal B, Llorente Ayala I, Juez Martel M, López Flores A. Carmona-Bayonas, Prognostic Nomogram and Patterns of Use of FOLFIRI-Aflibercept in Advanced Colorectal Cancer: A Real-World Data Analysis. *Oncologist*. 2019;24:e687–95.
 6. Deng J, Ren Z, Wen J, Wang B, Hou X, Xue Z, Chu X. Construction of a nomogram predicting the overall survival of patients with distantly metastatic non-small-cell lung cancer. *Cancer management research*. 2018;10:6143–56.
 7. Kim SY, Yoon MJ, Park YI, Kim MJ, Nam BH, Park SR. Nomograms predicting survival of patients with unresectable or metastatic gastric cancer who receive combination cytotoxic chemotherapy as first-line treatment. *Gastric cancer: official journal of the International Gastric Cancer Association the Japanese Gastric Cancer Association*. 2018;21:453–63.
 8. Doll KM, Rademaker A, Sosa JA. Practical Guide to Surgical Data Sets: Surveillance, Epidemiology, and End Results (SEER) Database, *JAMA surgery*, 153 (2018) 588–589.
 9. Shaikh WR, Weinstock MA, Halpern AC, Oliveria SA, Geller AC, Dusza SW. The characterization and potential impact of melanoma cases with unknown thickness in the United States' Surveillance, Epidemiology, and End Results Program, 1989–2008, *Cancer epidemiology*, 37 (2013) 64–70.
 10. Mo S, Cai X, Zhou Z, Li Y, Hu X, Ma X, Zhang L, Cai S, Peng J. Nomograms for predicting specific distant metastatic sites and overall survival of colorectal cancer patients: A large population-based real-world study. *Clinical translational medicine*. 2020;10:169–81.
 11. Taieb J, Kourie HR, Emile JF, Le Malicot K, Balogoun R, Tabernero J, Mini E, Folprecht G, Van Laethem JL, Mulot C, Bouché O, Aparicio T, Michel P, Thaler J, Bridgewater J, Van Cutsem E, Perkins G, Lepage C, Salazar R. P. Laurent-Puig, Association of Prognostic Value of Primary Tumor Location in Stage III Colon Cancer With RAS and BRAF Mutational Status. *JAMA oncology*. 2018;4:e173695.
 12. García-Alfonso P, Torres G, García G, Gallego I, Ortega L, Sandoval C, Muñoz A, Lloansí A. FOLFOXIRI plus biologics in advanced colorectal cancer. *Expert Opin Biol Ther*. 2019;19:411–22.
 13. Gresham G, Renouf DJ, Chan M, Kennecke HF, Lim HJ, Brown C, Cheung WY. Association between palliative resection of the primary tumor and overall survival in a population-based cohort of metastatic colorectal cancer patients. *Ann Surg Oncol*. 2014;21:3917–23.
 14. Faron M, Pignon JP, Malka D, Bourredjem A, Douillard JY, Adenis A, Elias D, Bouché O, Ducreux M, Is primary tumour resection associated with survival improvement in patients with colorectal cancer and unresectable synchronous metastases? A pooled analysis of individual data from four randomised trials, *European journal of cancer* (Oxford, England: 1990), 51 (2015) 166–176.
 15. Turner N, Tran B, Tran PV, Sinnathamby M, Wong HL, Jones I, Croxford M, Desai J, Tie J, Field KM, Kosmider S, Bae S, Gibbs P. Primary Tumor Resection in Patients With Metastatic Colorectal Cancer

- Is Associated With Reversal of Systemic Inflammation and Improved Survival, *Clinical colorectal cancer*, 14 (2015) 185–191.
16. Ishihara S, Hayama T, Yamada H, Nozawa K, Matsuda K, Miyata H, Yoneyama S, Tanaka T, Tanaka J, Kiyomatsu T, Kawai K, Nozawa H, Kanazawa T, Kazama S, Yamaguchi H, Sunami E, Kitayama J, Hashiguchi Y, Sugihara K, Watanabe T. Prognostic impact of primary tumor resection and lymph node dissection in stage IV colorectal cancer with unresectable metastasis: a propensity score analysis in a multicenter retrospective study. *Ann Surg Oncol*. 2014;21:2949–55.
 17. Cirocchi R, Trastulli S, Abraha I, Vettoretto N, Boselli C, Montedori A, Parisi A, Noya G, Platell C, Non-resection versus resection for an asymptomatic primary tumour in patients with unresectable stage IV colorectal cancer, *The Cochrane database of systematic reviews*, (2012) Cd008997.
 18. Wong SF, Wong HL, Field KM, Kosmider S, Tie J, Wong R, Tacey M, Shapiro J, Nott L, Richardson G, Cooray P, Jones I, Croxford M, Gibbs P. Primary Tumor Resection and Overall Survival in Patients With Metastatic Colorectal Cancer Treated With Palliative Intent, *Clinical colorectal cancer*, 15 (2016) e125-132.
 19. Yun JA, Huh JW, Park YA, Cho YB, Yun SH, Kim HC, Lee WY, Chun HK. The role of palliative resection for asymptomatic primary tumor in patients with unresectable stage IV colorectal cancer. *Dis Colon Rectum*. 2014;57:1049–58.
 20. Huang Y, Zhao M, Yin J, Lu T, Yang X, Yuan G, Li M, Liu Y, Zhan C, Wang Q. Pulmonary metastasis in newly diagnosed colon-rectal cancer: a population-based nomogram study. *Int J Colorectal Dis*. 2019;34:867–78.
 21. Lieu CH, Renfro LA, de Gramont A, Meyers JP, Maughan TS, Seymour MT, Saltz L, Goldberg RM, Sargent DJ, Eckhardt SG, Eng C. Association of age with survival in patients with metastatic colorectal cancer: analysis from the ARCAD Clinical Trials Program. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*. 2014;32:2975–84.
 22. Yan Q, Zhang K, Guo K, Liu S, Wasan HS, Jin H, Yuan L, Feng G, Shen F, Shen M, Ma S, Ruan S. Value of tumor size as a prognostic factor in metastatic colorectal cancer patients after chemotherapy: a population-based study, *Future oncology (London, England)*, 15 (2019) 1745–1758.
 23. Garufi C, Giacomini E, Torsello A, Sperduti I, Melucci E, Mottolese M, Zeuli M, Ettorre GM, Ricciardi T, Cognetti F, Magnani M, Ruzzo A. Gender effects of single nucleotide polymorphisms and miRNAs targeting clock-genes in metastatic colorectal cancer patients (mCRC). *Sci Rep*. 2016;6:34006.

Figures

Figure 1

Flow chart of the study, including selection and exclusion criteria.

Figure 2

Kaplan-Meier curves of OS (A) and CSS (B) for patients with only bone, brain, liver, or lung metastasis (n = 14,360).

Figure 3

Nomograms for patients with CRC and OLM. (A) Nomogram for 1- and 3-year-associated OS probability. (B) Nomogram for 1- and 3-year-associated CSS survival probability.

Figure 4

ROC analysis for internal and external validation cohorts (A, B) The ROC of 1- and 3-year OS for internal validation; (C, D) the ROC of 1- and 3-year CSS for internal validation; (E, F) the ROC of 1- and 3-year OS for external validation; (G, H) the ROC of 1- and 3-year CSS for external validation.

Figure 5

Calibration plots of OS and CSS associated nomograms in both internal and external validation cohorts (A, B) Calibration plots of 1- and 3-year OS for internal validation; (C, D) calibration plots of 1- and 3-year CSS for internal validation; (E, F) calibration plots of 1- and 3-year OS for external validation; (G, H) calibration plots of 1- and 3-year CSS for external validation.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarymaterials.docx](#)