Proceedings of Anticancer Research, 2024, Volume 8, Issue 6
http://ojs.bbwpublisher.com/index.php/PAR
ISSN Online: 2208-3553
ISSN Print: 2208-3545

Establishment and Validation of Diagnostic and Prognostic Prediction Models for Liver Metastasis in Patients with Rectal Cancer: A SEER-based Study

Huimin Wang[1,2,3], Ya Zheng[2,3], Zhaofeng Chen[2,3]*

[1]The First Clinical Medical College, Lanzhou University, Lanzhou, China
[2]Department of Gastroenterology, The First Hospital of Lanzhou University, Lanzhou, China
[3]Gansu Province Clinical Research Center for Digestive Diseases, The First Hospital of Lanzhou University, Lanzhou, China

***Corresponding author:** Zhaofeng Chen, zhfchen@lzu.edu.cn

Abstract: *Background:* Rectal cancer is one of the most common gastrointestinal tumors, among which the liver is the most common site of distant metastasis and liver metastasis that leads to poor prognosis. This study aimed to develop and validate a diagnostic nomogram to predict the occurrence of rectal cancer with liver metastasis (RCLM) and a prognostic nomogram to predict cancer-specific survival (CSS) in RCML patients. *Methods:* Data on patients with rectal cancer diagnosed between 2010 and 2013 were collected from the Surveillance, Epidemiology, and End Results (SEER) database. Univariate logistic regression and multivariate logistic regression were used to determine the independent risk factors of RCLM. Univariate Cox proportional hazards regression and multivariate Cox proportional hazards regression were used to identify independent prognostic factors for RCLM. This study then developed two novel nomograms, and the results were evaluated by receiver operating characteristic (ROC) curves, calibration curves, and decision curve analysis (DCA). *Results:* A total of 29,367 patients with rectal cancer were included. Among them, 3,403 patients (11.59%) had liver metastases at the time of diagnosis. The independent risk factors of RCLM included sex, race, tumor grade, AJCC N, CEA, marital status, tumor size, total number of primary tumors, and histological type. Age, chemotherapy, total number of primary tumors, surgery sites, and histological type were independent prognostic factors of patients with RCLM. The results of ROC curves, calibration curves, and DCA in the development, validation, and testing sets confirmed that the diagnostic nomogram can precisely predict the occurrence of RCLM. The results of ROC curves, calibration curves, DCA, C-indexes, and Kaplan–Meier (K-M) survival curves in the development, validation, and testing sets confirmed that the prognostic nomogram could precisely predict the prognosis of RCLM. *Conclusion:* The two nomograms are expected to be effective tools for predicting the risk of liver metastasis for patients with rectal cancer and personalized prognosis prediction for patients with RCLM, which may benefit clinical decision-making.

Keywords: Rectal cancer; Liver metastasis; Nomogram; Predictor; Diagnosis; Prognosis

Online publication: November 26, 2024

1. Introduction

Rectal cancer is the eighth most common malignant tumor in the world, with a high mortality rate of about 340,000 lives yearly [1,2]. It has become one of the significant public health problems threatening human health. However, colorectal cancer was often seen as a whole cohort to analyze the risk factors and prognostic factors in many studies [3-5]. There is evidence that colon cancer and rectal cancer differ in incidence, risk factors, mortality, patterns of distant metastasis, and clinical prognosis, and the incidence of rectal cancer is higher than that of colon cancer [6-9]. Therefore, further study of patients with rectal cancer as the independent subgroup is warranted. Among all distant metastases of rectal cancer, the incidence of liver metastasis was 42%, and the liver is the most common site of distant metastasis [10,11]. Patients with rectal cancer often have a poor prognosis due to liver metastasis and whose 5-year survival rate is less than 50% [12,13]. Chemotherapy, radiotherapy, and surgery all have different prognostic effects on liver metastasis of rectal cancer [14-16]. The rates of local failure and distant metastasis are substantial in these patients, even after undergoing aggressive treatments including resection of primary and metastatic liver tumors [17-19]. In addition, as far as the researchers know, people use a variety of machine learning methods. For example, Qiu includes seven clinical characteristics and determines the prediction model of liver metastases of rectal cancer to predict the occurrence of this disease [20]. However, no specific nomogram has been established for RCLM patients to further clarify and verify the risk or prognosis of RCLM. To address this limitation, this study integrated the latest large sample with comprehensive clinical information from the SEER database, including more clinicopathological features and information on chemotherapy, radiotherapy, and surgery, and developed two novel nomograms to be applied to the diagnosis and prognosis of RCLM.

In recent years, nomogram has been widely used to evaluate the diagnosis and prognosis of cancer patients, because it is convenient and accurate, and it is a good choice for us to study the disease [21,22]. Thus, this study identified a representative cohort from the Surveillance, Epidemiology, and End Results (SEER) database to evaluate the incidence, risk factors, and prognosis of RCLM, and developed two nomograms for predicting the risk of liver metastasis in rectal cancer patients and CSS of patients with RCLM, respectively.

2. Materials and methods

2.1. Study design and cohort selection

The Surveillance, Epidemiology, and End Results (SEER) database was an open-access cancer database covering around 30% of the United States (US) population, which recorded information about cancer incidence, treatment, and survival [23]. The data of patients with rectal cancer were collected from the SEER database using SEER*Stat software (version 8.4.3; https://seer.cancer.gov/seerstat/). Patients with rectal cancer from 2010 to 2013 were identified using "site record ICD-O-3/WHO 2008 = 'Rectum'" and "Year of diagnosis = '2010, 2011, 2012, 2013'" from the database "Incidence — SEER Research Data, 17 Registries, Nov 2022 Sub." The included variables were age, sex, race, grade, AJCC T, AJCC N, CEA, marital status, tumor size, total number of primary tumors, and histological type. Exclusion criteria: (1) Age less than 18 years old or more than 90 years old; (2) Patients diagnosed by autopsy or death certificate; (3) No information available. Finally, 29,367 patients diagnosed with rectal cancer were included in this study, including 3,403 patients with liver metastases. All patients were used to form a diagnostic cohort to explore risk factors of RCLM and to form a diagnostic prediction nomogram. In addition, among 3,403 patients with liver metastases, 3,359 patients had

definitive survival outcomes and were used to form a prognostic cohort. In this cohort, in addition to the above 11 variables, DX-bone (bone metastasis), DX-brain (brain metastasis), DX-lung (lung metastasis), radiotherapy or not, chemotherapy or not, and surgical site were also included. This study investigated prognostic factors for RCLM in this cohort and developed a novel prognostic nomogram.

In the diagnostic cohort, patients were randomly divided into the development (70%), and validation sets (30%) with a ratio of 7:3. As for the prognostic cohort, patients in the development and validation sets were composed of the patients who had liver metastasis from the diagnostic cohort with a ratio of 7:3. For each cohort, patients in the development set were used to construct the nomogram, and corresponding patients in the validation set were used to validate it. The data on rectal cancer and RCLM from 2014 to 2015 were extracted from the SEER database to form two independent testing sets for further validation of the two nomograms. The flow chart for patient screening is shown in **Figure 1**.

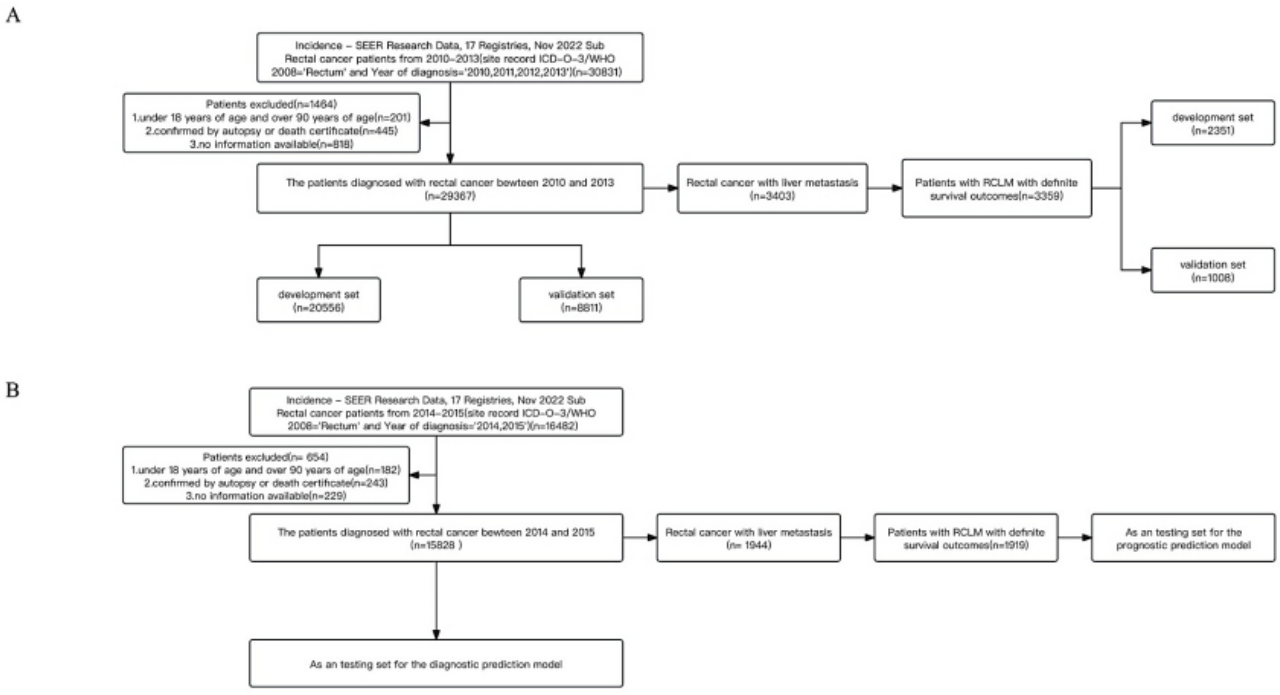

Figure 1. Flowchart of the patient screening: both A and B are population-based studies from the SEER database. The patient screening of the development sets and validation sets of the diagnostic prediction model and the prognostic prediction model **(A)**. The patient screening of the test sets of the diagnostic prediction model and the prognostic prediction model **(B)**

2.2. Survival outcome

In the diagnostic cohort, the primary outcome was the presence or absence of liver metastases. In the prognostic cohort, CSS was the primary outcome, which was defined as the time interval between the day diagnosed with rectal cancer and the day of death for rectal cancer.

2.3. Statistical analysis

In the present study, all statistical analysis was performed with Python 3.12.0. All patients (29,367 cases) with rectal cancer were randomly divided into the development and validation sets in Python software. The chi-

square test was used to analyze the distribution of the two data sets. In the diagnostic cohort, univariate logistic regression ($P < 0.05$) was used to determine the risk factors significantly related to liver metastasis of rectal cancer. The screened variables were included in the multivariate logistic analysis ($P < 0.05$) to determine the independent risk factors of patients with RCLM. In addition, the study constructed a new diagnostic nomogram based on independent risk factors. The time-dependent receiver operating characteristic (ROC) curves of the nomogram and all independent variables were generated, and the corresponding area under the curves (AUCs) was calculated to assess the discrimination. The decision curve analysis (DCA) of the nomogram and all independent variables were generated to assess the clinical effect. Moreover, the calibration curves were used to evaluate the performance of the nomogram.

In terms of prognostic factors, univariate Cox regression analysis was used to determine the related factors of CSS of patients with RCLM, and then significant variables with $P < 0.05$ were included in multivariate Cox analysis for further study. The K-M survival curves based on significant variables with $P < 0.05$ in the multivariate Cox analysis were constructed to determine the independent prognostic factors. A forest plot based on independent prognostic variables was drawn to analyze the effect of each subgroup on the CSS in patients with RCLM. A prognostic nomogram based on independent prognostic factors was constructed to predict the CSS of patients with RCLM, and the individual risk score was calculated using the formula of the nomogram. In addition, time-dependent ROC curves of the nomogram and all independent prognostic variables at 1.5, and 2.5 years were generated, and the corresponding time-dependent AUCs were applied to show the discrimination. DCA of nomogram and all independent prognostic variables at 1.5, and 2.5 years were generated to assess the clinical effect. Calibration curves of 1.5, and 2.5 years were plotted to evaluate the nomogram. According to the median risk score, patients with RCLM in the development, validation, and testing sets were divided into high-, medium- and low-risk groups. K–M survival curves and the log-rank test were performed to show the different CSS statuses between the three groups.

3. Results

3.1. Baseline characteristics of the study population

A total of 29,367 rectal cancer patients were recorded, and 20,556 and 8,811 patients were assigned to the development and validation sets. As shown in **Table 1**, the most common age was 60–90 years (57.39% in the development set and 56.36% in the validation set), the most common gender was male (58.68% in the development set and 59.24% in the validation set), and the most common race was white (77.27% in the development set and 77.04% in the validation set). The most common tumor grade was grade II (56.14% in the development set and 56.32% in the validation set). The most common AJCC T stage and AJCC N stage were T3 (38.39% in the development set and 38.75% in the validation set) and N0 (63.37% in the development set and 38.39% in the validation set), respectively. The most common number of primary tumors was single (73.73% in the development set and 73.95% in the validation set), the most common histological classification was adenocarcinoma (82.71% in the development set, 82.67% in the validation set), and the most common marital status was married (54.73% in the development set and 82.67% in the validation set). The most common tumor size was less than 988mm (69.50% in the development set and 69.48% in the validation set). Meanwhile, the chi-square test confirmed that the samples in the development and the internal validation sets were completely randomized (**Table 1**).

Table 1. Baseline characteristics of patients with rectal cancer [*n* (%)]

Variable	Overall (*n* = 29,367)	Development set (*n* = 20,556)	Validation set (*n* = 8,811)	χ^2	*P*
Age				2.645	0.104
18–59	12,603 (42.92%)	8,758 (42.61%)	3,845 (43.64%)		
60–90	16,764 (57.08%)	11,798 (57.39%)	4,966 (56.36%)		
Sex				0.777	0.378
Female	12,084 (41.15%)	8,493 (41.32%)	3,591 (40.76%)		
Male	17,283 (58.85%)	12,063 (58.68%)	5,220 (59.24%)		
Race				1.820	0.611
White	22,672 (77.20%)	15,884 (77.27%)	6,788 (77.04%)		
Black	3,222 (10.97%)	2,228 (10.84%)	994 (11.28%)		
Others	3,221 (10.97%)	2,262 (11.00%)	959 (10.88%)		
Unknown	252 (0.86%)	182 (0.89%)	70 (0.79%)		
Tumor grade				0.337	0.987
I	3,288 (11.20%)	2,303 (11.20%)	985 (11.18%)		
I	16,486 (56.14%)	11,524 (56.06%)	4,962 (56.32%)		
III	3,017 (10.27%)	2,109 (10.26%)	908 (10.31%)		
IV	396 (1.35%)	280 (1.36%)	116 (1.32%)		
Unknown	6,180 (21.04%)	4,340 (21.11%)	1,840 (20.88%)		
AJCC T stage				14.897	0.011
Tis	1,096 (3.73%)	795 (3.87%)	301 (3.42%)		
T1	7,378 (25.12%)	5,123 (24.92%)	2,255 (25.59%)		
T2	3,608 (12.29%)	2,476 (12.05%)	1,132 (12.85%)		
T3	11,306 (38.50%)	7,892 (38.39%)	3,414 (38.75%)		
T4	2,275 (7.75%)	1,644 (8.00%)	631 (7.16%)		
Unknown	3,704 (12.61%)	2,626 (12.77%)	1,078 (12.23%)		
AJCC N stage				1.188	0.756
N0	18,651 (63.51%)	13,027 (63.37%)	5,624 (63.83%)		
N1	7,200 (24.52%)	5,044 (24.54%)	2,156 (24.47%)		
N2	2,087 (7.11%)	1,481 (7.20%)	606 (6.88%)		
Unknown	1,429 (4.87%)	1,004 (4.88%)	425 (4.82%)		
CEA				2.262	0.520
Negative	7,592 (25.85%)	5,365 (26.10%)	2,227 (25.28%)		
Positive	7,168 (24.41%)	5,008 (24.36%)	2,160 (24.51%)		
Borderline	74 (0.25%)	51 (0.25%)	23 (0.26%)		
Unknown	14,533 (49.49%)	10,132 (49.29%)	4,401 (49.95%)		

Table 1 (Continued)

Variable	Overall (*n* = 29,367)	Development set (*n* = 20,556)	Validation set (*n* = 8,811)	χ^2	*P*
Marital status				6.183	0.045
Unmarried	10,967 (37.34%)	7,771 (37.80%)	3,196 (36.27%)		
Married	16,194 (55.14%)	11,251 (54.73%)	4,943 (56.10%)		
Unknown	2,206 (7.51%)	1,534 (7.46%)	672 (7.63%)		
Tumor size				1.604	0.448
< 988 mm	20,409 (69.50%)	14,287 (69.50%)	6,122 (69.48%)		
≥ 988 mm	250 (0.85%)	184 (0.90%)	66 (0.75%)		
Unknown	8,708 (29.65%)	6,085 (29.60%)	2,623 (29.77%)		
Total number				0.154	0.695
Single	21,671 (73.79%)	15,155 (73.73%)	6,516 (73.95%)		
Multiple	7,696 (26.21%)	5,401 (26.27%)	2,295 (26.05%)		
Histological type				2.462	0.782
Adenocarcinoma	24,285 (82.69%)	17,001 (82.71%)	7,284 (82.67%)		
Squamous carcinoma	673 (2.29%)	477 (2.32%)	196 (2.22%)		
Signet-ring cell carcinoma	200 (0.68%)	132 (0.64%)	68 (0.77%)		
Neuroendocrine carcinoma	603 (2.05%)	424 (2.06%)	179 (2.03%)		
Carcinoid	2,933 (9.99%)	2,060 (10.02%)	873 (9.91%)		
Others	673 (2.29%)	462 (2.25%)	211 (2.39%)		

3.2. Incidence and risk factors of RCLM

A total of 3,403 cases (11.59%) confirmed as liver metastases at the initial diagnosis and 25,964 cases (88.41%) without it. As shown in **Table 2**, the study performed univariate logistic regression ($P < 0.05$) on 11 potential factors, and the results showed 9 related variables, including sex, race, tumor grade, AJCC N, CEA, marital status, tumor size, total number of primary tumor, and histological type. In addition, multivariate logistic regression ($P < 0.05$) showed that sex, race, tumor grade, AJCC N, CEA, marital status, tumor size, total number of primary tumors, and histological type were independent risk predictors for liver metastasis of patients with rectal cancer (**Table 2**).

Table 2. Univariate and multivariate logistic regression analysis of patients with rectal cancer in the development set

Variable	Univariate analysis			Multivariate analysis		
	OR	95% *CI*	*P*	*OR*	95% *CI*	*P*
Age						
18–59	Reference					
60–90	0.855	0.731–1.001	0.052			

Table 2 (Continued)

Variable	Univariate analysis			Multivariate analysis		
	OR	95% *CI*	*P*	*OR*	95% *CI*	*P*
Sex						
Female	Reference			Reference		
Male	1.439	1.224–1.694	< 0.001	1.439	1.229–1.686	< 0.001
Race						
White	Reference			Reference		
Black	1.102	0.860–1.403	0.435	1.118	0.879–1.412	0.356
Others	0.603	0.455–0.790	< 0.001	0.621	0.472–0.806	< 0.001
Unknown	0.204	0.030–0.782	0.045	0.236	0.036–0.861	0.061
Tumor grade						
I	Reference			Reference		
II	1.569	1.114–2.252	0.012	1.477	1.061–2.096	0.025
III	2.264	1.543–3.367	< 0.001	2.244	1.550–3.296	< 0.001
IV	3.206	1.712–5.891	< 0.001	2.516	1.367–4.539	0.0025
Unknown	1.886	1.305–2.774	< 0.001	1.805	1.264–2.620	0.0015
AJCC T stage						
Tis	Reference					
T1	Inf	Inf	0.942			
T2	Inf	Inf	0.946			
T3	Inf	Inf	0.943			
T4	Inf	Inf	0.942			
Unknown	Inf	Inf	0.937			
AJCC N stage						
N0	Reference			Reference		
N1	2.32	1.924–2.800	< 0.001	2.246	1.891–2.669	< 0.001
N2	2.457	1.858–3.234	< 0.001	2.371	1.820–3.072	< 0.001
Unknown	3.018	2.269–4.009	< 0.001	6.571	5.089–8.482	< 0.001
CEA						
Negative	Reference			Reference		
Positive	8.852	6.876–11.545	< 0.001	10.118	7.893–13.142	< 0.001
Borderline	0	Inf	0.986	0	Inf	0.949
Unknown	2.467	1.889–3.257	< 0.001	2.525	1.941–3.321	< 0.001
Marital status						
Unmarried	Reference			Reference		
Married	0.814	0.694–0.956	0.012	0.761	0.652–0.889	< 0.001
Unknown	0.681	0.468–0.971	0.039	0.663	0.462–0.933	0.022

Table 2 (Continued)

Variable	Univariate analysis			Multivariate analysis		
	OR	95% *CI*	*P*	*OR*	95% *CI*	*P*
Tumor size						
< 988 mm	Reference			Reference		
≥ 988 mm	0.698	0.211–1.913	0.515	0.838	0.257–2.251	0.746
Unknown	1.442	1.209–1.719	< 0.001	1.932	1.640–2.274	< 0.001
Total number						
Single	Reference			Reference		
Multiple	0.537	0.438–0.654	< 0.001	0.534	0.439–0.645	< 0.001
Histological type						
Adenocarcinoma	Reference			Reference		
Squamous carcinoma	0.597	0.338–1,001	0.061	0.827	0.481–1.350	0.469
Signet-ring cell carcinoma	0.154	0.048–0.409	< 0.001	0.236	0.076–0.603	0.0056
Neuroendocrine carcinoma	1.564	0.941–2.540	0.076	2.714	1.661–4.332	< 0.001
Carcinoid	0.03	0.009–0.074	< 0.001	0.071	0.022–0.169	< 0.001
Others	1.388	0.942–2.031	0.094	2.15	1.483–3.091	< 0.001

3.3. Diagnostic nomogram development and validation

A novel nomogram for predicting the risk of liver metastases in patients with rectal cancer was established based on the nine independent predictors (**Figure 2A**). Then, the ROC curves of the development and validation sets were established, and the corresponding AUCs of the nomogram in the development and validation sets were 0.834 and 0.814, respectively (**Figure 2B–E**). The DCA curves of the development and validation sets (**Figure 2C–F**) showed that the nomogram had a good clinical effect.

Meanwhile, the ROC (**Figure 3A** and **3B**) and DCA curves (**Figure 3C** and **3D**) for all independent predictors in the development and validation sets were generated. The results showed that the new nomogram had better discrimination power and clinical effect than any individual factor, indicating that the diagnostic nomogram can be used as an accurate tool for the diagnosis of liver metastasis in patients with rectal cancer. More importantly, the calibration curves of the nomogram in the development and validation sets showed good consistency between the observed and predicted results (**Figure 2D–G**).

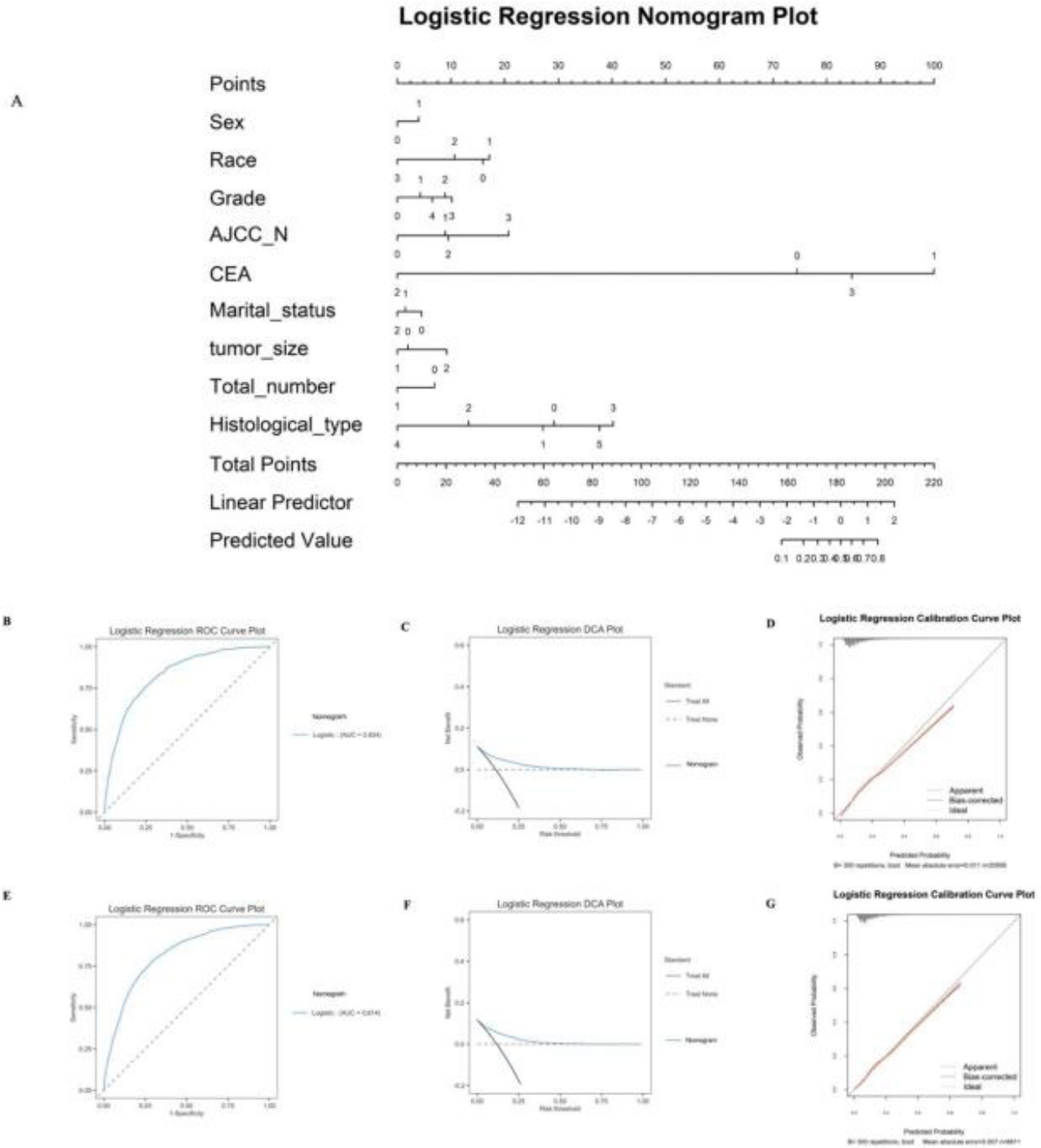

Figure 2. Construction and validation of a diagnostic nomogram. A nomogram to estimate the risk of liver metastasis in patients with rectal cancer **(A)**. The receiver operating characteristic curve **(B)**, decision curve analysis **(C)**, and calibration curve **(D)** of the development set, and the receiver operating characteristic curve **(E)**, decision curve analysis **(F)**, and calibration curve **(G)** of the validation set

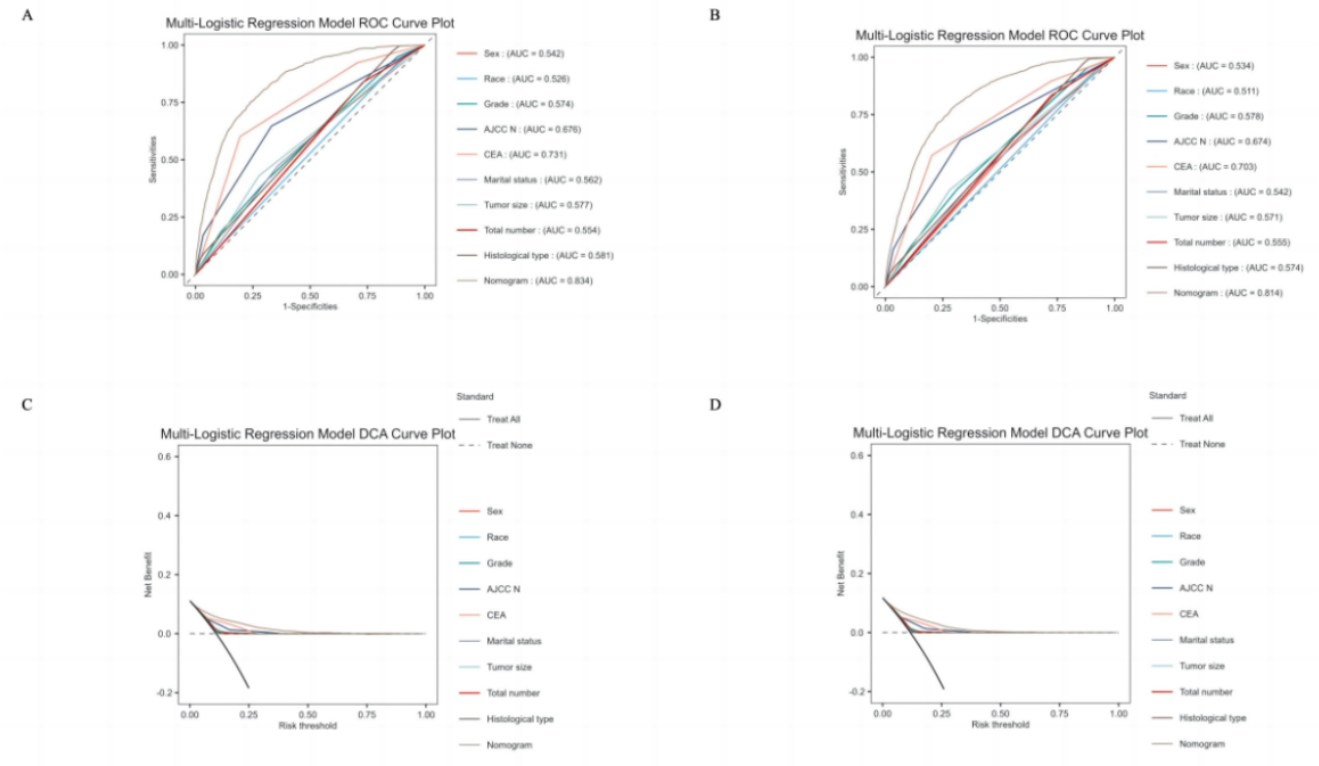

Figure 3. Comparison of area under the receiver operating characteristic curves between the diagnostic nomogram and all independent risk factors, including sex, race, tumor grade, AJCC N, CEA, marital status, tumor size, total number of primary tumors, and histological type in the development set **(A)** and the validation set **(B)**.Comparison of DCA curves between the diagnostic nomogram and all independent risk factors, including sex, race, tumor grade, AJCC N, CEA, marital status, tumor size, total number of primary tumors, and histological type in the development set **(C)** and the validation set **(D)**

3.4. Prognostic factors for patients with RCLM

In the present study, 3,359 eligible patients with RCLM were used to explore prognostic factors. As shown in **Table 3**, 1,096 patients (32.63%) received surgery, 1,128 (33.58%) received radiotherapy, and 2,434 (72.46%) received chemotherapy, and the chi-square test showed that all variables were not significantly different between the development set and the validation set, confirming that data between the development set and the validation set are randomly assigned. Univariate and multivariate Cox regression analyses were performed on the development set to evaluate each prognostic factor (**Table 4**), age, AJCC T, AJCC N, radiotherapy, chemotherapy, tumor size, DX-bone, DX-lung, total number, surgery sites, histological type were significantly ($P < 0.05$) identified in univariate analysis in the development set. The further Cox regression analysis also showed that age, chemotherapy, total number, histological type, and surgery sites were independent prognostic factors for CSS (**Figure 4**), which were included in the nomogram. Subsequently, a forest plot based on independent prognostic factors was used to determine the effect of each subgroup on the CSS. The researchers can see from the forest plot that: age > 60 years old was a risk factor for CSS compared with age < 60 years old, chemotherapy was a protective factor for CSS compared with no chemotherapy, carcinoid was a protective factor, signet ring cell carcinoma was a risk factor and other histological types have no significant significance

for CSS compared with adenocarcinoma, no surgery was a risk factor and surgery on other sites have no significant significance for CSS compared with surgery only on the primary site, and multiple tumors are a risk factor for CSS compared with single rectal cancer.

Table 3. Baseline clinical characteristics of patients diagnosed with RCLM

Variable	Overall (*n* = 3,359)	Development set (*n* = 2,351)	Validation set (*n* = 1008)	χ^2	*P*
Age				0.589	0.443
18–59	1,544 (45.97%)	1,070 (45.51%)	474 (47.02%)		
60–90	1,815 (54.03%)	1,281 (54.49%)	534 (52.98%)		
Sex				0.100	0.752
Female	1,168 (34.77%)	822 (34.96%)	346 (34.33%)		
Male	2,191 (65.23%)	1,529 (65.04%)	662 (65.67%)		
Race				0.401	0.940
White	2,663 (79.28%)	1,863 (79.24%)	800 (79.37%)		
Black	377 (11.22%)	264 (11.23%)	113 (11.21%)		
Others	317 (9.44%)	223 (9.49%)	94 (9.33%)		
Unknown	2 (0.06%)	1 (0.04%)	1 (0.10%)		
Tumor grade				5.365	0.252
I	160 (4.76%)	125 (5.32%)	35 (3.47%)		
I	1,784 (53.11%)	1,240 (52.74%)	544 (53.97%)		
III	517 (15.39%)	359 (15.27%)	158 (15.67%)		
IV	72 (2.14%)	51 (2.17%)	21 (2.08%)		
Unknown	826 (24.59%)	576 (24.50%)	250 (24.80%)		
AJCC T stage				2.558	0.634
T1	474 (14.11%)	342 (14.55%)	132 (13.10%)		
T2	137 (4.08%)	92 (3.91%)	45 (4.46%)		
T3	1,194 (35.55%)	840 (35.73%)	354 (35.12%)		
T4	446 (13.28%)	303 (12.89%)	143 (14.19%)		
Unknown	1,108 (32.99%)	774 (32.92%)	334 (33.13%)		
AJCC N stage				1.832	0.608
N0	1,203 (35.81%)	825 (35.09%)	378 (37.50%)		
N1	1,214 (36.14%)	859 (36.54%)	355 (35.22%)		
N2	389 (11.58%)	277 (11.78%)	112 (11.11%)		
Unknown	553 (16.46%)	390 (16.59%)	163 (16.17%)		
Radiotherapy				0.915	0.339
No	2,231 (66.42%)	1,549 (65.89%)	682 (67.66%)		
Yes	1,128 (33.58%)	802 (34.11%)	326 (32.34%)		

Table 3 (Continued)

Variable	Overall (*n* = 3,359)	Development set (*n* = 2,351)	Validation set (*n* = 1008)	χ^2	*P*
Chemotherapy				1.376	0.241
No	925 (27.54%)	633 (26.92%)	292 (28.97%)		
Yes	2,434 (72.46%)	1,718 (73.08%)	716 (71.03%)		
CEA				2.622	0.454
Negative	324 (9.65%)	226 (9.61%)	98 (9.72%)		
Positive	1,950 (58.05%)	1,356 (57.68%)	594 (58.93%)		
Borderline	3 (0.09%)	1 (0.04%)	2 (0.20%)		
Unknown	1,082 (32.21%)	768 (32.67%)	314 (31.15%)		
Marital status				0.272	0.873
Unmarried	1,508 (44.89%)	1,050 (44.66%)	458 (45.44%)		
Married	1,680 (50.01%)	1179 (50.15%)	501 (49.70%)		
Unknown	171 (5.09%)	122 (5.19%)	49 (4.86%)		
Tumor size				1.023	0.600
< 988 mm	1,919 (57.13%)	1,354 (57.59%)	565 (56.05%)		
≥ 988 mm	19 (0.57%)	12 (0.51%)	7 (0.69%)		
Unknown	1,421 (42.30%)	985 (41.90%)	436 (43.25%)		
Bone metastases				1.854	0.396
No	2,992 (89.07%)	2,083 (88.60%)	909 (90.18%)		
Yes	245 (7.29%)	178 (7.57%)	67 (6.65%)		
Unknown	122 (3.63%)	90 (3.83%)	32 (3.17%)		
Lung metastases				2.512	0.285
No	2,237 (66.60%)	1,585 (67.42%)	652 (64.68%)		
Yes	1,003 (29.86%)	683 (29.05%)	320 (31.75%)		
Unknown	119 (3.54%)	83 (3.53%)	36 (3.57%)		
Brain metastases				3.307	0.191
No	3,185 (94.82%)	2,219 (94.39%)	966 (95.83%)		
Yes	30 (0.89%)	24 (1.02%)	6 (0.60%)		
Unknown	144 (4.29%)	108 (4.59%)	36 (3.57%)		
Total number				0.820	0.365
Single	2,801 (83.39%)	1,951 (82.99%)	850 (84.33%)		
Multiple	558 (16.61%)	400 (17.01%)	158 (15.67%)		
Surgery sites				7.438	0.114
Only primary site	691 (20.57%)	501 (21.31%)	190 (18.85%)		
Only other sites	102 (3.04%)	74 (3.15%)	28 (2.78%)		
Primary and other sites	303 (9.02%)	211 (8.97%)	92 (9.13%)		

Table 3 (Continued)

Variable	Overall (*n* = 3,359)	Development set (*n* = 2,351)	Validation set (*n* = 1008)	χ^2	*P*
No surgery	2,238 (66.63%)	1,543 (65.63%)	695 (68.95%)		
Unknown	25 (0.74%)	22 (0.94%)	3 (0.30%)		
Histological type				3.513	0.621
Adenocarcinoma	2,961 (88.15%)	2,061 (87.66%)	900 (89.29%)		
Squamous carcinoma	63 (1.88%)	43 (1.83%)	20 (1.98%)		
Signet-ring cell carcinoma	28 (0.83%)	21 (0.89%)	7 (0.69%)		
Neuroendocrine carcinoma	92 (2.74%)	65 (2.76%)	27 (2.68%)		
Carcinoid	21 (0.63%)	17 (0.72%)	4 (0.40%)		
Others	194 (5.78%)	144 (6.13%)	50 (4.96%)		

Table 4. Univariate and multivariate Cox regression analysis predicting CSS in patients with RCLM in the development set

Variable	Univariate analysis			Multivariate analysis		
	HR	95% *CL*	*P*	*HR*	95% *CL*	*P*
Age						
18–59	Reference			Reference		
60–90	1.617	1.285–2.037	0	1.383	1.078–1.775	0.011
Sex						
Female	Reference					
Male	1.201	0.946–1.525	0.133			
Race						
White	Reference					
Black	0.801	0.534–1.202	0.284			
Others	0.843	0.569–1.251	0.397			
Unknown	0	Inf	0.994			
Tumor grade						
I	Reference					
I	1.016	0.617–1.673	0.951			
III	1.235	0.705–2.163	0.462			
IV	1.053	0.48–2.313	0.897			
Unknown	1.207	0.7–2.08	0.499			
AJCC T stage						
T1	Reference			Reference		
T2	1.046	0.619–1.768	0.867	1.33	0.761–2.326	0.317
T3	0.993	0.691–1.426	0.969	1.271	0.861–1.878	0.228

Table 4 (Continued)

Variable	Univariate analysis			Multivariate analysis		
	HR	95% *CL*	*P*	*HR*	95% *CL*	*P*
T4	1.4	0.897–2.184	0.139	1.377	0.863–2.199	0.179
Unknown	1.538	1.044–2.267	0.029	1.261	0.825–1.929	0.284
AJCC N stage						
N0	Reference			Reference		
N1	0.674	0.518–0.877	0.003	0.723	0.544–0.962	0.056
N2	0.883	0.615–1.267	0.500	1.191	0.805–1.764	0.382
Unknown	1.304	0.895–1.902	0.167	0.934	0.608–1.436	0.757
Radiotherapy						
No	Reference			Reference		
Yes	0.687	0.543–0.87	0.002	0.895	0.686–1.169	0.416
Chemotherapy						
No	Reference			Reference		
Yes	0.397	0.303–0.519	0	0.363	0.267–0.493	0
CEA						
Negative	Reference					
Positive	1.023	0.732–1.43	0.893			
Borderline	0	Inf	0.993			
Unknown	1.388	0.977–1.971	0.067			
Marital status						
Unmarried	Reference					
Married	0.838	0.658–1.067	0.152			
Unknown	1.453	0.91–2.318	0.117			
Tumor size						
< 988 mm	Reference			Reference		
≥ 988 mm	0	Inf	0.992	0	Inf	0.993
Unknown	1.356	1.065–1.725	0.013	1.028	0.774–1.366	0.848
Bone metastases						
No	Reference			Reference		
Yes	1.882	1.125–3.148	0.016	1.632	0.957–2.784	0.072
Unknown	2.203	1.198–4.051	0.011	1.748	0.753–4.056	0.194
Lung metastases						
No	Reference			Reference		
Yes	1.481	1.082–2.028	0.014	1.331	0.954–1.857	0.092
Unknown	1.646	0.769–3.525	0.2	0.888	0.319–2.473	0.821

Table 4 (Continued)

Variable	Univariate analysis			Multivariate analysis		
	HR	95% *CL*	*P*	*HR*	95% *CL*	*P*
Brain metastases						
No	Reference					
Yes	1.227	0.171–8.792	0.839			
Unknown	1.569	0.878–2.804	0.128			
Total number						
Single	Reference			Reference		
Multiple	1.712	1.316–2.228	0	1.621	1.221–2.151	0.001
Surgery sites						
Only primary site	Reference			Reference		
Only other sites	0.862	0.399–1.864	0.707	1.096	0.488–2.457	0.825
Primary and other sites	0.934	0.67–1.303	0.688	1.198	0.847–1.693	0.307
No surgery	1.469	1.115–1.936	0.006	1.496	1.069–2.091	0.019
Unknown	2.314	0.846–6.328	0.102	1.487	0.524–4.216	0.456
Histological type						
Adenocarcinoma	Reference			Reference		
Squamous carcinoma	1.761	0.933–3.323	0.081	1.212	0.607–2.42	0.587
Signet-ring cell carcinoma	4.707	1.735–12.768	0.002	4.163	1.512–11.46	0.006
Neuroendocrine carcinoma	0.77	0.362–1.637	0.49	0.631	0.282–1.412	0.263
Carcinoid	0.705	0.29–1.717	0.442	0.151	0.056–0.407	0
Others	1.527	0.905–2.577	0.113	1.339	0.778–2.305	0.292

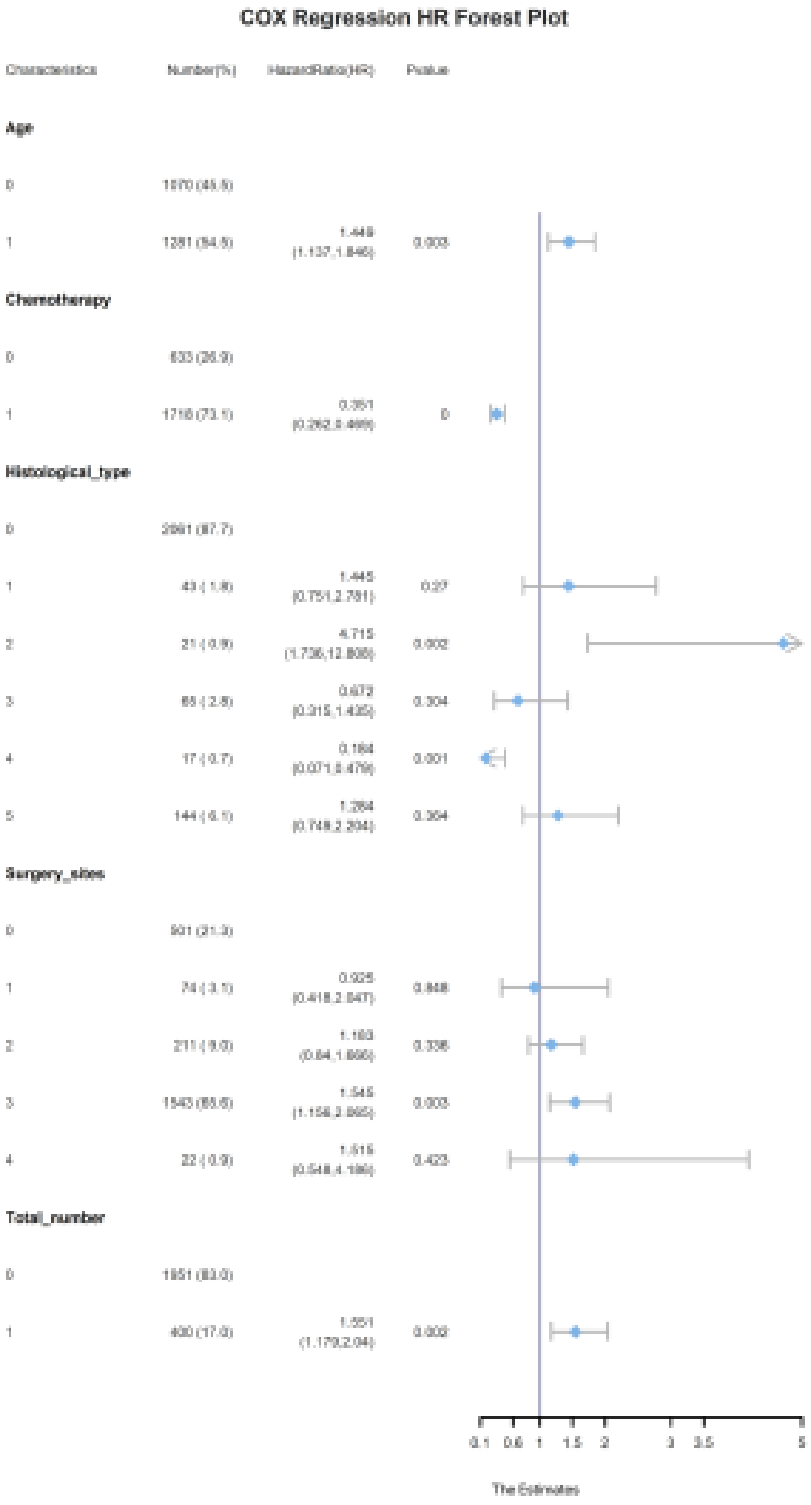

Figure 4. A forest plot of meaningful parameters in the multivariate Cox regression analysis. (Age: 0 represents 18 to 60 years old, and 1 represents 60 to 90 years old. Chemotherapy: 0 represents no, and 1 represents yes. Total number: 0 represents a single tumor, and 1 represents multiple tumors. Histological type: 0 represents adenocarcinoma, 1 represents squamous cell carcinoma, 2 represents signet ring cell carcinoma, 3 represents neuroendocrine carcinoma, 4 represents carcinoid, and 5 represents other types. Surgery sites: 0 represents surgery at the primary site only, 1 represents surgery at the non-primary site, 2 represents surgery at both the primary site and the non-primary site, 3 represents no surgery, and 4 represents other conditions.)

3.5. Prognostic nomogram development and validation

Based on the five independent prognostic factors, a prognostic nomogram was developed to predict the 1.5- and 2.5-year cancer-specific survival prediction of patients with RCLM (**Figure 5**). The C-indexes of the nomogram in the development set for the 1.5, and 2.5 years were 0.714 and 0.738, and 0.709 and 0.722 in the validation set. The ROC curves showed that the 1.5- and 2.5-year AUCs for the nomogram were 0.768 and 0.771 in the development set (**Figure 6A–D**), and 0.743 and 0.699 in the validation set (**Figure 7A–D**), indicating that the CSS prediction model had good predictive performance in both the development set and the validation set. The 1.5- and 2.5-year DCA curves in the development set (**Figure 6B–E**) and the validation set (**Figure 7B–E**) indicated that the nomogram performed well in clinical application. In addition, the 1.5- and 2.5-year calibration curves also showed uniformity between nomogram-predicted CSS and the actual outcome in the development set (**Figure 6C–F**) and validation set (**Figure 7C–F**). Furthermore, the study compared the 1.5- and 2.5-year ROC curves between the nomogram and five independent prognostic factors in the development set (**Figure 8A–C**) and the validation set (**Figure 8E–G**), and the 1.5- and 2.5-year DCA curves between the nomogram and five independent prognostic factors in the development set (**Figure 8B–D**) and the validation set (**Figure 8F–H**). The results revealed that the prognostic nomogram was superior to the five independent prognostic factors in predicting the 1.5- and 2.5-year cancer-specific survival prediction.

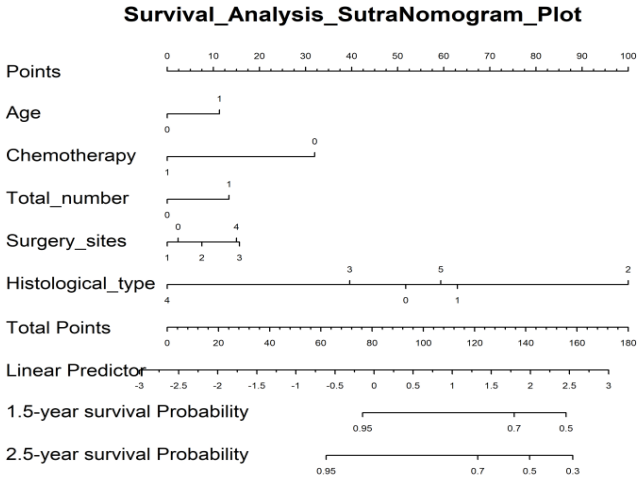

Figure 5. A prognostic nomogram for predicting the CSS of patients with RCLM for 1.5 and 2.5 years

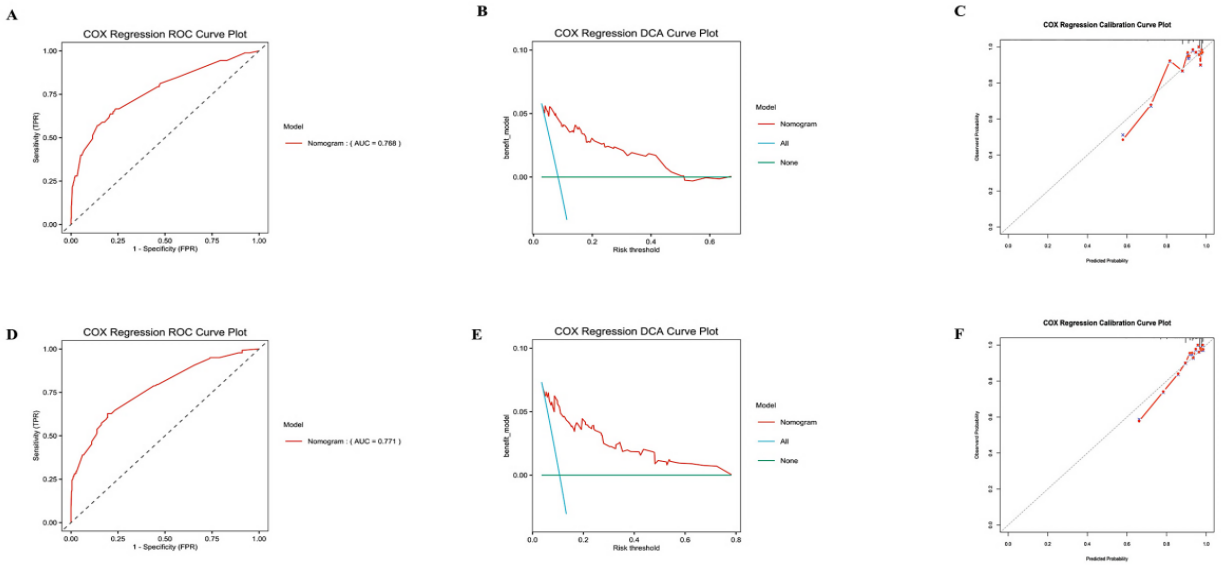

Figure 6. The receiver operating characteristic curve, the decision curves analysis, and the calibration curves of the prognostic nomogram for 1.5 **(A, B, C)**, and 2.5 years **(D, E, F)** in the development set

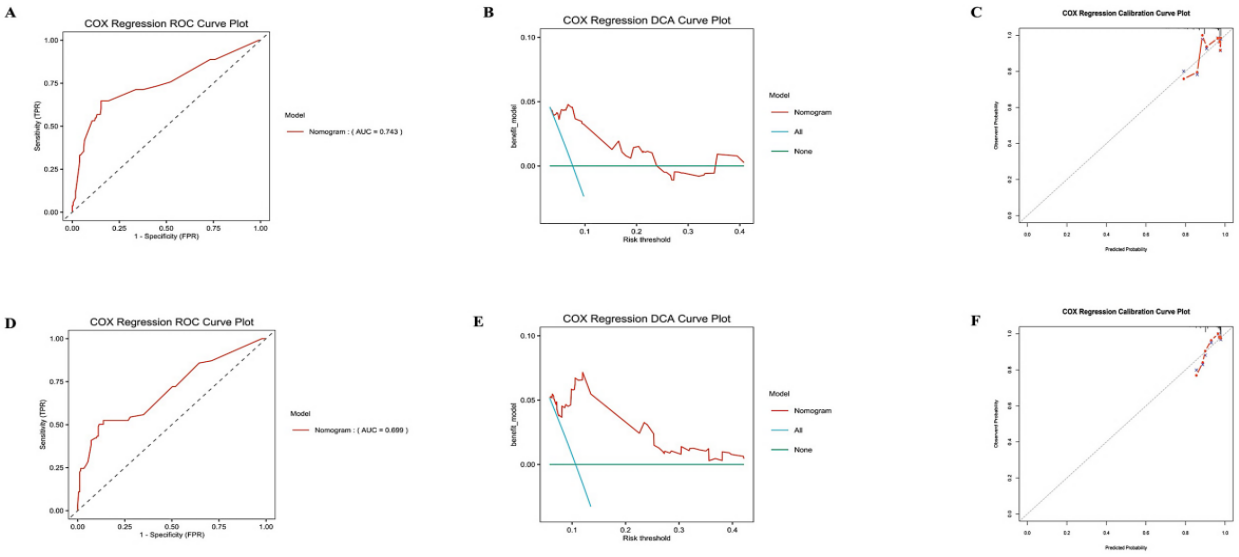

Figure 7. The receiver operating characteristic curve, the decision curves analysis, and the calibration curves of the prognostic nomogram for 1.5 **(A, B, C)**, and 2.5 years **(D, E, F)** in the validation set

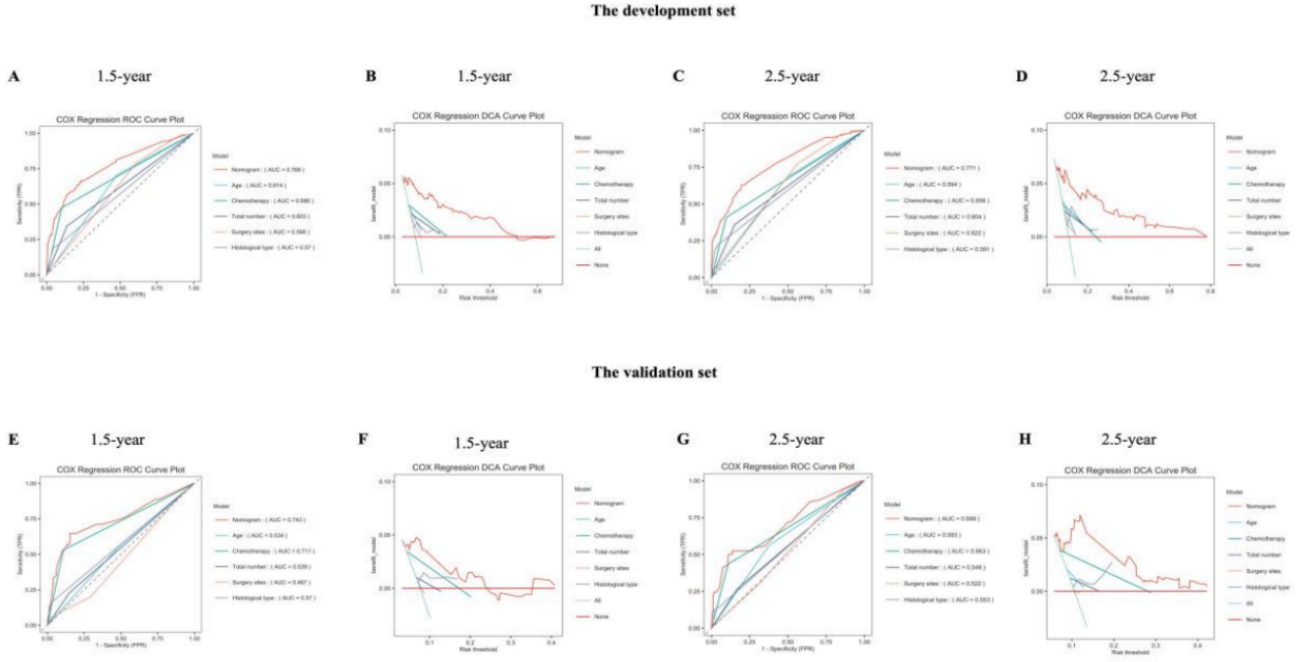

Figure 8. Comparison of area under the receiver operating characteristic curves between the prognostic nomogram and all independent factors, including age, chemotherapy, total number (from the primary tumors), surgery sites, and histological type for the 1.5, and 2.5 years in the development set **(A, C)** and the validation set **(E, G)**. Comparison of DCA curves between the prognostic nomogram and all independent factors, including age, chemotherapy, total number (from the primary tumors), surgery sites, histological type for the1.5, and 2.5 years in the development set **(B, D)**, and the validation set **(F, H)**.

3.6. Validating the diagnostic nomogram and the prognostic nomogram in the two independent testing sets, respectively

The basic information of the two independent testing sets from the SEER database is shown in **Table 5**. 15,828 patients were enrolled as the testing set of the diagnostic nomogram. In the testing set, the AUC of the ROC curve of the diagnostic nomogram was 0.897 (**Figure 9A**), and the DCA curve (**Figure 9B**), and calibration curve (**Figure 9C**) demonstrated the good performance of the diagnostic nomogram. In addition, the ROC curve and DCA curve of the diagnostic nomogram were above the curve of the single risk factor (**Figure 9D–E**), indicating that the diagnostic nomogram had a better diagnostic effect than the single risk factor in clinical practice. 1,919 patients with RCLM had a definitive survival outcome and formed the testing set for the prognostic nomogram. The 1.5-year and 2.5-year C-indexes of the prognostic nomogram in the testing set were 0.715 (95% CI, 0.662–0.744) and 0.716 (95%,0.668–0.749), respectively. The 1.5-year and 2.5-year AUCs were 0.737 and 0.734, respectively (**Figure 10A–F**). In addition, the DCA curve (**Figure 10B–G**) and the calibration curve (**Figure 10C–H**) indicated good performance of the prognostic nomogram. The 1.5-year and 2.5-year ROC curves and DCAs of the prognostic nomogram were all above the curve of each single prognostic factor (**Figure 10D–I, E–J**), indicating that the diagnostic prediction model had better discrimination and more ideal clinical benefit than the single prognostic factor.

Table 5. The basic information of the two independent testing sets from the SEER database

Variable	Rectal cancer patients (*n* = 15,828)	Rectal cancer patients with liver metastasis (*n* = 1,919)
Age		
18–59	7,078 (44.72%)	857 (44.66%)
60–90	8,750 (55.28%)	1,062 (55.34%)
Sex		
Female	6,466 (40.85%)	716 (37.31%)
Male	9,362 (59.15%)	1,203 (62.69%)
Race		
White	12,156 (76.80%)	1,498 (78.06%)
Black	1,680 (10.61%)	203 (10.58%)
Others	1,816 (11.47%)	214 (11.15%)
Unknown	176 (1.11%)	4 (0.21%)
Tumor grade		
I	2,355 (14.88%)	118 (6.15%)
I	8,725 (55.12%)	994 (51.80%)
III	1483 (9.37%)	265 (13.81%)
IV	219 (1.38%)	49 (2.55%)
Unknown	3,046 (19.24%)	493 (25.69%)
AJCC T stage		
Tis	509 (3.22%)	0 (0.00%)
T1	3,903 (24.66%)	273 (14.23%)
T2	1,723 (10.89%)	61 (3.18%)
T3	6,229 (39.35%)	661 (34.45%)
T4	1359 (8.59%)	276 (14.38%)
Unknown	2,105 (13.30%)	648 (33.77%)
AJCC N stage		
N0	9,452 (59.72%)	689 (35.90%)
N1	4,311 (27.24%)	763 (39.76%)
N2	1,266 (8.00%)	198 (10.32%)
Unknown	799 (5.05%)	269 (14.02%)
Radiotherapy		
No	8,361 (52.82%)	1,274 (66.39%)
Yes	7,467 (47.18%)	645 (33.61%)
Chemotherapy		
No	6,985 (44.13%)	481 (25.07%)
Yes	8,843 (55.87%)	1,438 (74.93%)

Table 5 (Continued)

Variable	Rectal cancer patients (*n* = 15,828)	Rectal cancer patients with liver metastasis (*n* = 1,919)
CEA		
Negative	4,211 (26.60%)	183 (9.54%)
Positive	4,025 (25.43%)	1,126 (58.68%)
Borderline	38 (0.24%)	3 (0.16%)
Unknown	7,554 (47.73%)	607 (31.63%)
Marital status		
Unmarried	5,976 (37.76%)	832 (43.36%)
Married	8,619 (54.45%)	997 (51.95%)
Unknown	1,233 (7.79%)	90 (4.69%)
Tumor size		
< 988 mm	11,468 (72.45%)	1,180 (61.49%)
≥ 988 mm	150 (0.95%)	11 (0.57%)
Unknown	4,210 (26.60%)	728 (37.94%)
Bone metastases		
No	15,534 (98.14%)	1,708 (89.00%)
Yes	237 (1.50%)	165 (8.60%)
Unknown	57 (0.36%)	46 (2.40%)
Lung metastases		
No	14,823 (93.65%)	1,269 (66.13%)
Yes	934 (5.90%)	609 (31.74%)
Unknown	71 (0.45%)	41 (2.14%)
Brain metastases		
No	15,727 (99.36%)	1,851 (96.46%)
Yes	42 (0.27%)	18 (0.94%)
Unknown	59 (0.37%)	50 (2.61%)
Total number		
Single	12,044 (76.09%)	1,596 (83.17%)
Multiple	3,784 (23.91%)	323 (16.83%)
Surgery sites		
Only primary site	10,752 (67.93%)	301 (15.69%)
Only other sites	135 (0.85%)	70 (3.65%)
Primary and other sites	589 (3.72%)	186 (9.69%)
No surgery	4,276 (27.02%)	1,355 (70.61%)
Unknown	76 (0.48%)	7 (0.36%)

Table 5 (Continued)

Variable	Rectal cancer patients (n = 15,828)	Rectal cancer patients with liver metastasis (n = 1,919)
Histological type		
Adenocarcinoma	12,820 (81.00%)	1,686 (87.86%)
Squamous carcinoma	411 (2.60%)	23 (1.20%)
Signet-ring cell carcinoma	94 (0.59%)	7 (0.36%)
Euroendocrine carcinoma	252 (1.59%)	59 (3.07%)
Carcinoid	1,867 (11.80%)	21 (1.09%)
Others	384 (2.43%)	123 (6.41%)

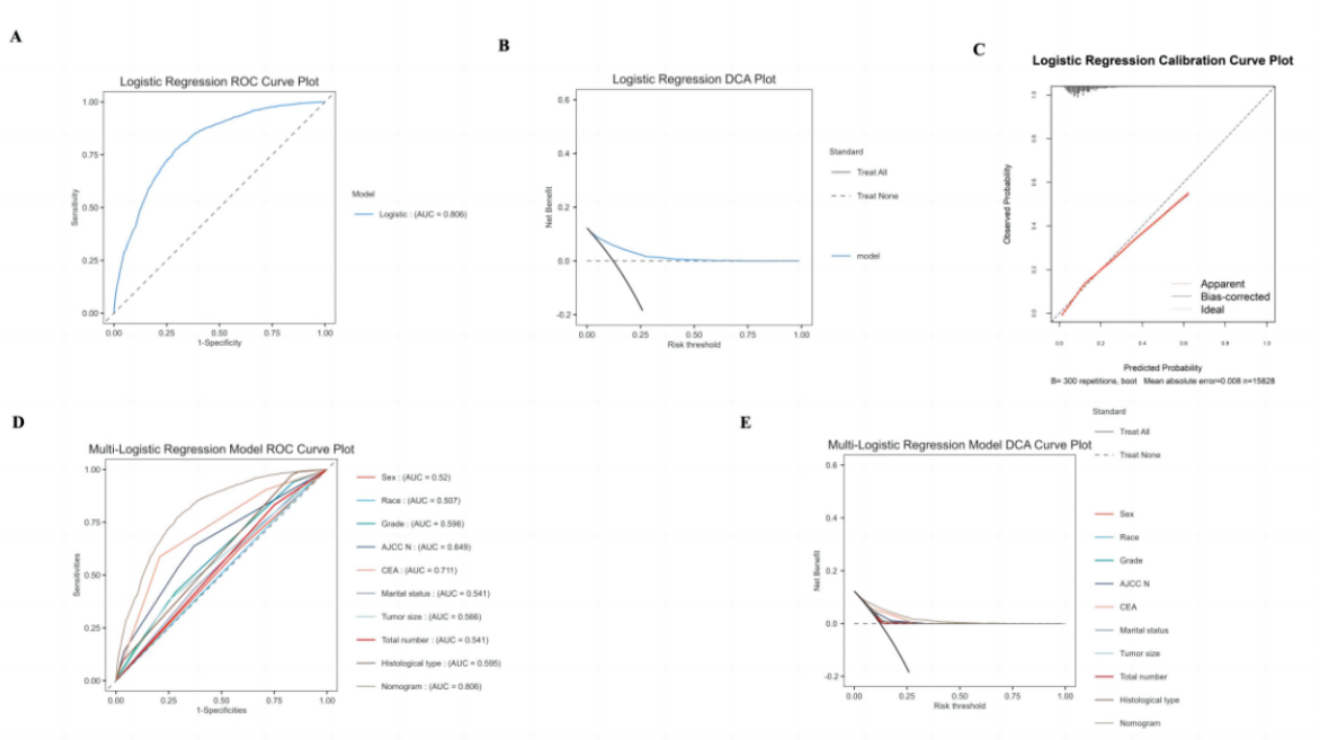

Figure 9. Validating the diagnostic nomogram in the testing set. Time-dependent ROC curve analysis **(A)**, the decision curve analysis **(B)**, and the calibration curve **(C)** of the diagnostic nomogram in the testing set. Comparison of area under the receiver operating characteristic curves between the nomogram and all independent factors **(D)**, including AJCC N, chemotherapy, CEA, DX-lung, and surgery sites in the validation set. Comparison of the decision curve analysis between the nomogram and all independent factors **(E)**, including AJCC N, chemotherapy, CEA, DX-lung, and surgery sites in the testing set

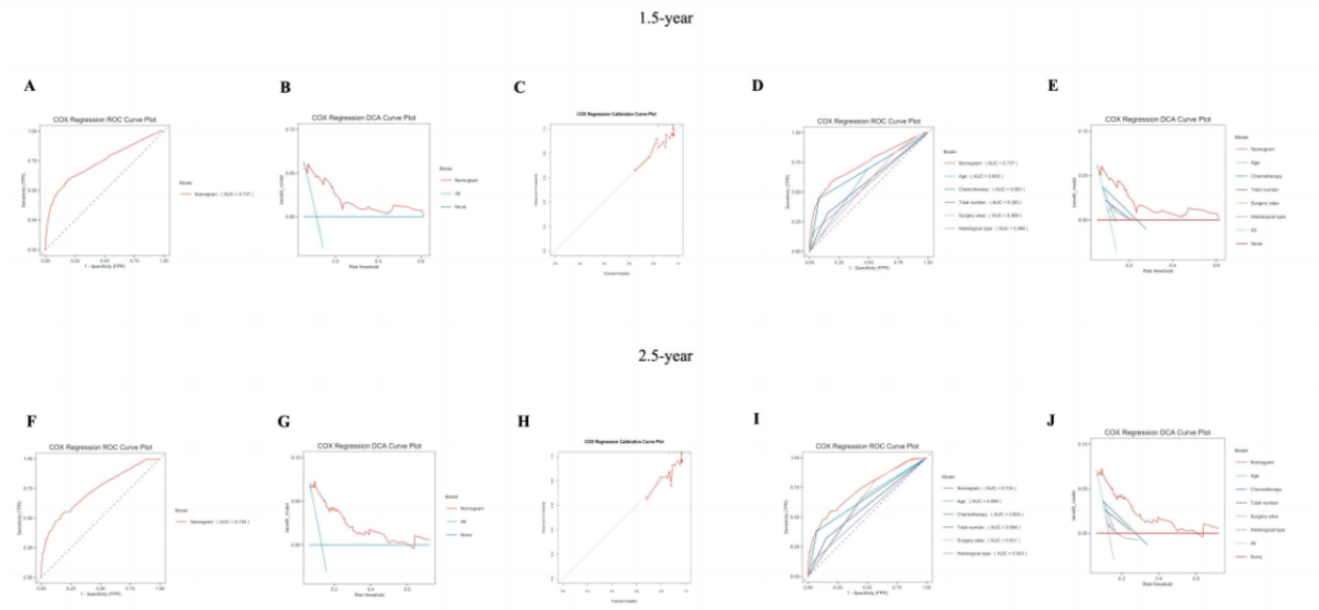

Figure 10. Validating the prognostic nomogram in the testing set. The receiver operating characteristic curve, the decision curves analysis, and the calibration curves of the prognostic nomogram for 1.5 **(A, B, C)**, and 2.5 years **(F, G, H)** in the testing set. Comparison of area under the receiver operating characteristic curves between the prognostic nomogram and all independent factors, including age, chemotherapy, total number, surgery sites, and histological type for the 1.5 **(D)**, and 2.5 years **(I)** in the testing set. Comparison of DCA curves between the prognostic nomogram and all independent factors, including age, chemotherapy, total number, surgery sites, and histological type for the 1.5 **(E)**, and 2.5 years **(J)** in the testing set

3.7. Survival analysis

According to the prognostic model established in this study, patients with RCLM can be divided into three groups: low risk, median risk, and high risk. The study calculated the risk score according to the constructed prognostic prediction model, and the patients were divided into low-risk, median-risk, and high-risk groups according to the risk score. The Kaplan–Meier survival curves of patients in different risk subgroups in the development set, the validation set, and the testing set showed worse CSS conditions in the high-risk group than in the median-risk and low-risk group ($P < 0.05$) (**Figure 11**). The results suggest that the model can be used to classify patients with RCLM into three groups with significantly different prognoses.

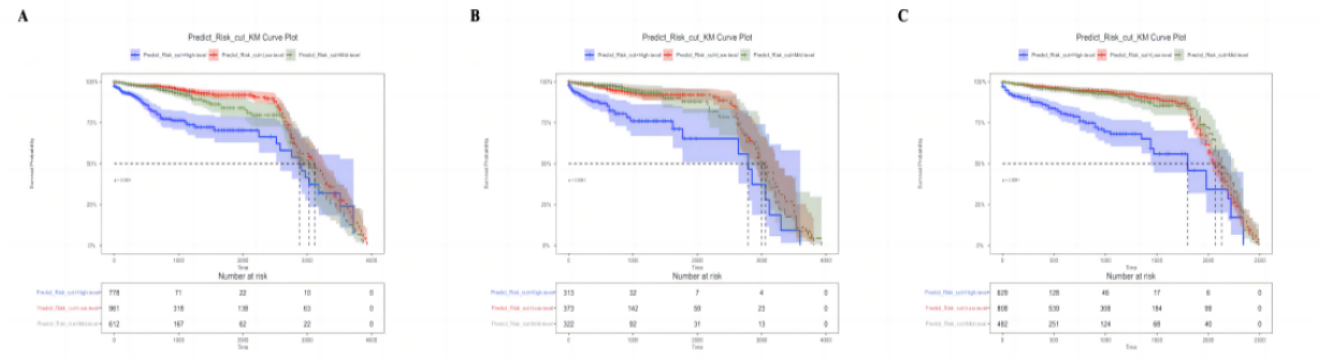

Figure 11. The Kaplan–Meier survival curves of patients in different risk subgroups in the development set **(A)**, the validation set **(B)**, and the testing set **(C)**

4. Discussion

Distant metastasis from rectal cancer usually results in poorer survival and quality of life, half of them had liver-limited disease (LLD) [24,25]. Therefore, researchers must identify the effective risk factors and prognostic factors in patients with liver metastases from rectal cancer, to facilitate early prevention and diagnosis, and effectively evaluate the prognosis of such patients. In the present study, the researchers constructed a diagnostic nomogram for predicting the occurrence of liver metastases in patients with rectal cancer and a prognostic nomogram for patients with RCLM. By obtaining the data of several key accessible variables on the nomograms, diagnosis-related and prognosis-related scores can be calculated, which can provide guidance for further clinical evaluation and intervention [26].

In recent years, more and more attention has been paid to RCLM, but most of them have been conducted on the molecular level and radiomics rather than clinicopathological features. Circulating sCD40, circulating exosomal miR-141-3p and miR-375, CD73, the miR-329-3p/Netrin-1-CD146 Complex has been confirmed to be related to liver metastasis of rectal cancer and can be used for early prediction of liver metastasis, thereby greatly improving the survival rate of rectal cancer patients [27–30]. In terms of radiomics, many studies also indicated that radiomics based on primary rectal cancer could provide a non-invasive way to predict the risk of rectal cancer with liver metastasis in clinical practice, such as radiomics obtained from MRI [31-33]. Researchers need to point out that these studies are often small in sample size and single-center studies, which lack adequate validation, making these biomarkers and partial radiomics not generalizable and difficult to apply immediately to clinical management. The study was not only based on the large sample size but also contained a large number of clinicopathological features and treatment information and it was found that the incidence of liver metastasis was 11.59%, which was lower than the 15% to 20% in the previous studies [11,34].

A novel diagnostic nomogram based on five independent predictors (sex, race, tumor grade, AJCC N, CEA, marital status, tumor size, total number of primary tumors, and histological type) was developed. A large tumor volume means a long growth cycle, leading to more proliferative and aggressive tumor cells, which increases liver metastasis [35]. Similarly, in the study, primary tumor sizes ≥ 988 mm (HR>1) were more likely to develop liver metastases than primary tumor sizes < 988 mm.

Black people (HR>1) are more prone to liver metastasis than white people, which may be caused by related genes or survival environment, and further human studies are needed to discuss. Married patients received more psychological and financial support and showed better adherence than unmarried people. Given the current research status, the study included marital status in the model to explore the relationship between marital status and liver metastases in patients with rectal cancer. The results also showed that married patients (HR < 1) were more likely to have a slower occurrence of liver metastases than unmarried patients, and the inclusion of marital status improved the stability and robustness of the model. Several researchers have found that men and patients with low-grade stage are more likely to develop liver metastases from rectal cancer, which is consistent with the results of this study [36,37].

Compared with women, men are more likely to develop liver metastases, and the researchers suspect that this is often related to bad lifestyle habits, such as smoking and drinking, which destroy liver cells so that the liver responds poorly to tumor invasion. Compared with extended rectal cancer, patients with stage I rectal cancer are still in the early stage, and in the face of cancer cell invasion, the liver can still make corresponding compensatory resistance and slow down the occurrence of liver metastasis. CEA is a standard tumor marker on colorectal cancer cell membranes and embryonic mucosal cells. Studies have shown that CEA is often

associated with distant metastasis of rectal cancer [38]. Therefore, it is not surprising that serum CEA levels in patients with colorectal cancer liver metastasis in this study can predict the occurrence of liver metastasis. Among the different histological types, the incidence of neuroendocrine tumors is 1000 cases per year; 11% are located in the gastrointestinal tract, and the prognosis is poor, most often in the esophagus and large intestine. In the study, rectal neuroendocrine tumors were more prone to liver metastasis than other histological types [39]. Several studies have shown that patients with regional lymph node metastases are more likely to develop liver metastases of rectal cancer [40,41]. The liver is one of the most abundant organs of lymphoid tissue in the body. Therefore, when local lymph nodes metastasize, tumors are more likely to metastasize. The study also found that a single primary rectal cancer tumor (HR > 1) was more likely to metastasize than multiple primary tumors. This prediction model fully integrates various risk factors that may affect liver metastasis of rectal cancer and has achieved excellent prediction performance. When verifying the diagnostic prediction model, the researchers found that the calibration curve, ROC curve, and DCA all showed good performance in the development set, validation set, and test set, with AUC greater than 0.7. The calibration curve of the diagnostic prediction model is close to the diagonal line, which indicates that the model has good predictability and accuracy. The ROC curves for all individual risk factors were under the curves of the constructed diagnostic model, indicating that the model was more predictable than any independent risk factor. DCA results also showed that the predictive model produced a higher clinical benefit than any single risk factor.

At present, only the prognostic factors of RCLM have been investigated, and no suitable prediction model has been constructed. Therefore, prediction models based on multiple prognostic factors are still lacking. To explore the prognosis of the unique subgroup of patients with RCLM, the researchers performed the identification of significant prognostic factors and the establishment of prediction models in this study, which could provide valuable guidance for the prognosis evaluation and individualized management of patients with RCLM. In this study, the researchers also developed a 3- and 5-year prognostic prediction model for RCLM based on a large sample from the SEER database. Five parameters (age, chemotherapy, total number, surgery sites, and histological type) significantly correlated with CSS in RCLM patients were used as independent prognostic factors in the model. The results of the analysis showed that the risk of cancer-specific mortality was increased by about 38% (HR, 1.383) in patients older than 60 years old compared with patients younger than 60 years old, indicating that the prognosis of patients with RCLM tends to be worse in aging patients. Patients with RCLM with multiple tumors have an increased cancer-specific mortality of about 62% (HR, 1.621) compared with patients with a single tumor. Therefore, avoiding the development of multiple tumors may improve the survival of such patients. Furthermore, compared with patients with RCLM whose histologic type was adenocarcinoma, cancer-specific mortality reduced by approximately 37% (HR, 0.631) for such patients with neuroendocrine carcinoma and by approximately 85% (HR, 0.151) for those with carcinoid, whereas cancer-specific mortality is relatively increased for other histologic types. So clinically, researchers can evaluate the prognosis of individual patients according to different histological types. In the research, “surgical sites” and “chemotherapy” were also used as significant prognostic predictors in the study. The study found that compared with patients who underwent surgery only on the primary site, patients who underwent surgery at other sites or no site had an increased risk of cancer-specific mortality (HR>1), indicating that patients who underwent surgery only on the primary site had a better prognosis. Chemotherapy-treated patients had about a 64% reduction (HR, 0.363) in cancer-specific mortality compared with patients who did not receive chemotherapy, indicating a better prognosis for chemotherapy-treated patients. There is evidence that rectal cancer with liver

metastasis is a challenging disease that requires chemotherapy, radiation, and surgery to optimize outcomes for individual patients [42]. According to the main international clinical guidelines, the recommended treatment for locally advanced rectal cancer is neoadjuvant chemoradiotherapy followed by surgery [43]. Compared to long-course chemoradiotherapy, total neoadjuvant treatment with short-course radiotherapy and chemotherapy significantly decreased the occurrence of metastases, particularly liver metastases. Among these treatments, surgical resection remains the principal curative approach that offers significant survival improvements [44]. Although patients with RCLM developed more lesions and some resistance to chemotherapy, continuous chemotherapy still plays an important role in prolonging life and provides a survival advantage for patients with RCLM, which highlights the urgent need for new treatment strategies, so some relevant clinical trials are still in progress [45,46]. In addition, the use of radiotherapy in metastatic rectal cancer is popular and also provides a certain survival space for the prognosis of patients with RCLM [47]. However, the results of the analysis excluded the parameter "radiotherapy," the researchers speculate that "radiotherapy" may have some collinearity with "chemotherapy." The researchers cannot exclude the importance of "radiotherapy" in the prognosis of patients with RCLM. In the validation of the prognostic prediction model, C-indexes, calibration curves, ROC curves, DCA curves, and K-M survival curves in the development set, the validation set, and the testing set all showed good predictive performance and clinical applicability.

The advantages of this study are as follows. First, the establishment of a nomogram for liver metastasis in patients with rectal cancer is very rare. A nomogram was used to show and apply the prediction models as a convenient form to predict various clinical outcomes, providing better guidance for RCLM-individualized medical judgment and decision-making. Secondly, when constructing the diagnostic prediction model, not only the logistic regression used for univariate analysis but also the AUCs of 18 independent variables were put into the univariate analysis to screen out more sensitive and accurate predictors. Multiple COX multivariate analyses were used to identify more accurate independent prognostic factors and to establish a prognostic predictive model. Third, this study included specific information about systemic therapy that other studies about RCLM did not. Finally, in the absence of external data, the study implemented more adequate verification tools and went back to the SEER database to verify the performance of the nomogram again.

However, the researchers should acknowledge some shortcomings of this study. First, partially missing or poorly informative data were excluded, exacerbating the risk of selection bias. Second, although the two predicting models were constructed in different development sets and validated in the validation sets and the testing sets, the same complete data were not available in any hospital or other databases for further validation, which may make prospective studies of patients with RCLM difficult. Third, the information collected in the SEER database was about the disease at the time of initial diagnosis, which meant that the hepatic metastases that occurred in the latter stage could not be included. In addition, although race did not affect the occurrence and prognosis of RCLM, most of our subjects were white, which made the applicability of the models to other ethnic groups unknown and required further study.

5. Conclusion

The study determined that AJCC N, chemotherapy, CEA, DX-lung, and surgical sites were the independent risk factors of liver metastasis for patients with rectal cancer, and age, chemotherapy, total number, surgery sites, histological type were the independent prognostic factors for the patients with RCLM. Two nomograms

could be used as an intuitive graphic tool for RCLM to quantitatively evaluate the risk and prognosis of liver metastasis in patients with rectal cancer, and guide the clinical decision-making.

Funding

(1) Department of Science and Technology of Gansu Province, Major Project of Joint Scientific Research Fund of Gansu Province "Gastric Cancer Epidemiology, Pathogenesis and Diagnosis and Treatment Technology Research" (Project No. 23JRRA1487).
(2) Lanzhou Science and Technology Bureau, the First Hospital of Lanzhou University Lanzhou R&D Platform Construction Project Lanzhou Science and Technology Program Digest "Screening and Verification of Early Screening Markers for Gastric Cancer and Precancerous Lesions" (Project No. 2023-1-19)
(3) The Foundation of The First Hospital of Lanzhou University, Hospital Foundation of the First Hospital of Lanzhou University Digestion 2021 "Study on the Mechanism of Exosome (LINC00957) Encoded Micropeptide Protein in Promoting the Occurrence and Development of Gastric Cancer"

Disclosure statement

The authors declare no conflict of interest.

Author contributions

Conceptualization: Huimin Wang, Ya Zheng, Zhaofeng Chen
Investigation: Huimin Wang
Software: Huimin Wang
Analysis: Huimin Wang
Writing – original draft: Huimin Wang
Writing – review & editing: Zhaofeng Chen and Ya Zheng

References

[1] Sung H, Ferlay J, Siegel RL, et al., 2021, Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians, 71(3): 209–249.
[2] Wilkinson N, 2020, Management of Rectal Cancer. The Surgical Clinics of North America, 100(3): 615–628.
[3] Wang Q, Shen K, Fei B, et al., 2023, Nomogram for Predicting Occurrence and Prognosis of Liver Metastasis in Elderly Colorectal Cancer Patients: A Population-based Study. Frontiers in Oncology, 2023(13): 1295650.
[4] Han L, Dai W, Mo S, et al., 2020, Nomogram to Predict the Risk and Survival of Synchronous Bone Metastasis in Colorectal Cancer: A Population-based real-world Analysis. International Journal of Colorectal Disease, 35(8): 1575–1585.
[5] Guo X, Liu Y, Liu LJ, et al., 2021, Development and Validation of Survival Nomograms in Colorectal Cancer Patients with Synchronous Liver Metastases Underwent Simultaneous Surgical Treatment of Primary and Metastatic Lesions. American Journal of Cancer Research, 11(6): 2654–2669.
[6] Sullivan DEO, Sutherland RL, Town S, et al., 2022, Risk Factors for Early-Onset Colorectal Cancer: A Systematic

Review and Meta-analysis. Clinical Gastroenterology and Hepatology, 20(6): 1229–1240.

[7] Rumpold H, Kirchweger P, Niedersu B, et al., 2021, Prognostic Value of Metastatic Pattern in Colorectal Cancer: A Multicenter Retrospective Analysis in a Real-life Cohort. Acta Oncologica, 60(2): 180–186.

[8] Hugen N, van de Velde CJH, de Wilt JHW, et al., 2014, Metastatic Pattern in Colorectal Cancer is Strongly Influenced by Histological Subtype. Annals of Oncology, 25(3): 651–657.

[9] Riihimaki M, Hemminki A, Sundquist J, et al., 2016, Patterns of Metastasis in Colon and Rectal Cancer. Scientific Reports, 2016(6): 29765.

[10] Mariathasan AB, Boye K, Dueland S, et al., 2021, Metastases in Locally Advanced Rectal Cancer Undergoing Curatively Intended Treatment. European Journal of Surgical Oncology, 47(9): 2377–2383.

[11] Besutti G, Damato A, Venturelli F, et al., 2021, Baseline Liver Steatosis has no Impact on Liver Metastases and Overall Survival in Rectal Cancer Patients. BMC Cancer, 21(1): 253.

[12] Chandra R, Karalis JD, Liu C, et al., 2021, The Colorectal Cancer Tumor Microenvironment and Its Impact on Liver and Lung Metastasis. Cancers (Basel), 13(24): 6206.

[13] Kawaguchi Y, Vauthey JN, 2020, The Landmark Series: Randomized Control Trials Examining Perioperative Chemotherapy and Postoperative Adjuvant Chemotherapy for Resectable Colorectal Liver Metastasis. Annals of Surgical Oncology, 27(11): 4263–4270.

[14] Liu Z, Meng X, Zhang H, et al., 2020, Predicting Distant Metastasis and Chemotherapy Benefit in Locally Advanced Rectal Cancer. Nature Communications, 11(1): 4308.

[15] Berardi R, Maccaroni E, Onofri A, et al., 2014, Locally Advanced Rectal Cancer: The Importance of a Multidisciplinary Approach. World Journal of Gastroenterology, 20(46): 17279–17287.

[16] Garcia-Aguilar J, Patil S, Gollub MJ, et al., 2022, Organ Preservation in Patients with Rectal Adenocarcinoma Treated with Total Neoadjuvant Therapy. Journal of Clinical Oncology, 40(23): 2546–2556.

[17] Kim JW, Kim YB, Kim NK, et al., 2010, The Role of Adjuvant Pelvic Radiotherapy in Rectal Cancer with Synchronous Liver Metastasis: A Retrospective Study. Radiation Oncology, 2010(5): 75.

[18] Akgul O, Cetinkaya E, Ersoz S, et al., 2014, Role of Surgery in Colorectal Cancer Liver Metastases. World Journal of Gastroenterology, 20(20): 6113–6122.

[19] Rompianesi G, Pegoraro F, Ceresa CD, et al., 2022, Artificial Intelligence in the Diagnosis and Management of Colorectal Cancer Liver Metastases. World Journal of Gastroenterology, 28(1): 108–122.

[20] Qiu B, Su XH, Qin X, et al., 2022, Application of Machine Learning Techniques in Real-world Research to Predict the Risk of Liver Metastasis in Rectal Cancer. Frontiers in Oncology, 2022(12): 1065468.

[21] Balachandran VP, Gonen M, Smith JJ, et al., 2015, Nomograms in Oncology: More than Meets the Eye. The Lancet (Oncology), 16(4): 173–180.

[22] Song C, Yu D, Wang Y, et al., 2020, Dual Primary Cancer Patients with Lung Cancer as a Second Primary Malignancy: A Population-Based Study. Frontiers in Oncology. 2020(10): 515606.

[23] Cronin KA, Ries LA, Edwards BK, 2014, The Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute. Cancer, 120(23): 3755–3757.

[24] Qiu B, Shen Z, Wu S, et al., 2023, A Machine Learning-based Model for Predicting Distant Metastasis in Patients with Rectal Cancer. Frontiers in Oncology, 2023(13): 1235121.

[25] Gelsomino F, Spallanzani A, Garajova I, 2019, The Treatment of Rectal Cancer with Synchronous Liver Metastases: A Matter of Strategy. Critical Reviews in Oncology/Hematology, 2019(139): 91–95.

[26] Chen B, Zeng Y, Liu B, et al., 2021, Risk Factors, Prognostic Factors, and Nomograms for Distant Metastasis in Patients

with Newly Diagnosed Osteosarcoma: A Population-Based Study. Frontiers in Endocrinology (Lausanne), 2021(12): 672024.

[27] Meltzer S, Torgunrud A, Abrahamsson H, et al., 2021, The Circulating Soluble Form of the CD40 Costimulatory Immune Checkpoint Receptor and Liver Metastasis Risk in Rectal Cancer. British Journal of Cancer, 125(2): 240–246.

[28] Meltzer S, Bjornetro T, Lyckander LG, et al., 2019, Circulating Exosomal miR-141-3p and miR-375 in Metastatic Progression of Rectal Cancer. Translational Oncology, 12(8): 1038–1044.

[29] Messaoudi N, Cousineau I, Arslanian E, et al., 2020, Prognostic Value of CD73 Expression in Resected Colorectal Cancer Liver Metastasis. Oncoimmunology. 9(1): 1746138.

[30] Fang X, Xu Y, Li K, et al., 2022, Exosomal lncRNA PCAT1 Promotes Tumor Circulating Cell-Mediated Colorectal Cancer Liver Metastasis by Regulating the Activity of the miR-329-3p/Netrin-1-CD146 Complex. Journal of Immunology Research, 2022: 9916228.

[31] Shu Z, Fang S, Ding Z, et al., 2019, MRI-based Radiomics Nomogram to Detect Primary Rectal Cancer with Synchronous Liver Metastases. Scientific Reports, 9(1): 3374.

[32] Liang M, Cai Z, Zhang H, et al., 2019, Machine Learning-based Analysis of Rectal Cancer MRI Radiomics for Prediction of Metachronous Liver Metastasis. Academic Radiology, 26(11): 1495–1504.

[33] Zhou N, Guo X, Sun H, et al., 2021, The Value of (18)F-FDG PET/CT and Abdominal PET/MRI as a One-Stop Protocol in Patients with Potentially Resectable Colorectal Liver Metastases. Frontiers in Oncology, 2021(11): 714948.

[34] PelvEx Collaborative, 2020, Management Strategies for Patients with Advanced Rectal Cancer and Liver Metastases using Modified Delphi Methodology: Results from the PelvEx Collaborative. Colorectal Disease, 22(9): 1184–1188.

[35] Bai BL, Wu ZY, Weng SJ, et al., 2023, Application of Interpretable Machine Learning Algorithms to Predict Distant Metastasis in Osteosarcoma. Cancer Medicine, 12(4): 5025–5034.

[36] Tang M, Wang H, Cao Y, et al., 2021, Nomogram for Predicting Occurrence and Prognosis of Liver Metastasis in Colorectal Cancer: A Population-based Study. International Journal of Colorectal Disease, 36(2): 271–282.

[37] Kow AWC, 2019, Hepatic Metastasis from Colorectal Cancer. Euroasian Journal of Hepato-gastroenterology, 10(6): 1274–1298.

[38] Sharma V, Sharma A, Raina V, et al., 2021, Metastatic Colorectal Cancer: Real Life Experience from an Indian Tertiary Care Center. BMC Cancer, 21(1): 630.

[39] Bruera G, Giuliani A, Romano L, et al., 2019, Poorly Differentiated Neuroendocrine Rectal Carcinoma with Uncommon Immune-histochemical Features and Clinical Presentation with a Subcutaneous Metastasis, Treated with First Line Intensive Triplet Chemotherapy plus Bevacizumab FIr-B/FOx Regimen: An Experience of Multidisciplinary Management in Clinical Practice. BMC Cancer, 19(1): 960.

[40] Liu H, Cui Y, Shen W, et al., 2016, Pretreatment Magnetic Resonance Imaging of Regional Lymph Nodes with Carcinoembryonic Antigen in Prediction of Synchronous Distant Metastasis in Patients with Rectal Cancer. Oncotarget, 7(19): 27199–27207.

[41] Liu LL, Sun JD, Xiang ZL, 2021, A Nomogram to Predict the Prognosis of Patients with Unresected Rectal Adenocarcinoma Undergoing Chemoradiotherapy: A Population-based Study. Journal of Cancer, 12(16): 4745–4761.

[42] Ganesh K, Wu C, O'Rourke KP, et al., 2019, A Rectal Cancer Organoid Platform to Study Individual Responses to Chemoradiation. Nature Medicine, 25(10): 1607–1614.

[43] Lopez-Campos F, Martin-Martin M, Fornell-Perez R, et al., 2020, Watch and Wait Approach in Rectal Cancer: Current Controversies and Future Directions. World Journal of Gastroenterology, 26(29): 4218–4239.

[44] Saad AM, Abdel-Rahman O, 2019, Initial Systemic Chemotherapeutic and Targeted Therapy Strategies for the

Treatment of Colorectal Cancer Patients with Liver Metastases. Expert Opinion on Pharmacotherapy, 20(14): 1767–1775.

[45] Kryeziu K, Moosavi SH, Bergsland CH, et al., 2021, Increased Sensitivity to SMAC Mimetic LCL161 Identified by Longitudinal ex vivo Pharmacogenomics of Recurrent, KRAS Mutated Rectal Cancer Liver Metastases. Journal of Translational Medicine, 19(1): 384.

[46] Song Y, Long J, Su X, et al., 2023, Case Report: Genetic and Immune Microenvironmental Characteristics of a Rectal Cancer Patient with MSS/PD-L1-Negative Recurrent Hepatopulmonary Metastasis who Achieved Complete Remission after Treatment with PD-1 Inhibitor. Frontiers in Immunology, 2023(14): 1197543.

[47] Moslim MA, Bastawrous AL, Jeyarajah DR, 2021, Neoadjuvant Pelvic Radiotherapy in the Management of Rectal Cancer with Synchronous Liver Metastases: Is It Worth It? Journal of Gastrointestinal Surgery, 25(9): 2411–2422.