OPEN ACCESS

A model established using marital status and other factors from the Surveillance, Epidemiology, and End Results database for early stage gastric cancer

Lixiang Zhang, Baichuan Zhou, Panquan Luo, Aman Xu, Wenxiu Han, Zhijian Wei

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/jim-2021-002285).

General Surgery, First Affiliated Hospital of Anhui Medical University, Hefei, Anhui, China

Correspondence to
Professor Aman Xu and Dr Zhijian Wei, General Surgery, First Affiliated Hospital of Anhui Medical University, Hefei 230022, Anhui, China; xuaman@ahmu.edu.cn, 305801533@qq.com

LZ and BZ contributed equally.

LZ and BZ are joint first authors.

Accepted 13 April 2022
Published Online First 5 July 2022

To cite: Zhang L, Zhou B, Luo P, *et al*. *J Investig Med* 2022;**70**:1373–1380.

ABSTRACT

Currently, the postoperative prognosis of early stage gastric cancer (GC) is difficult to accurately predict. In particular, social factors are not frequently used in the prognostic assessment of early stage GC. Therefore, this study aimed to combine the clinical indicators and social factors to establish a predictive model for early stage GC based on a new scoring system. A total of 3647 patients with early stage GC from the Surveillance, Epidemiology, and End Results database were included in this study. A Kaplan-Meier survival analysis was used to compare differences in prognosis between different marital status, as an innovative prognostic indicator. Univariate and multivariate analyses were used to screen available prediction factors and then build a nomogram using the Cox proportional hazard regression model. The univariate analysis and multivariate analysis revealed that age at diagnosis, sex, histology, stage_T, surgery, tumor size, and marital status were independent prognostic factors of overall survival. Both the C-index and calibration curves confirmed that the nomogram had a great predictive effect on patient prognosis in training and testing sets. This nomogram based on clinical indicators and marital status can effectively help patients with early stage GC in the future.

Significance of this study

What is already known about this subject?

⇒ The prognosis of early stage gastric cancer (GC) has always been the focus of GC research.

⇒ According to National Comprehensive Cancer Network guidelines, the prognosis of patients with early stage GC is correlated with age, tumor site, pathological stage, and other factors, but social factors were not taken into account and the impact of these factors on prognosis has not been quantified and comprehensively applied.

⇒ The prognostic evaluation ability of marital status has been fully recognized for patients with liver cancer, lung cancer, and other tumors.

What are the new findings?

⇒ Based on previous studies, we innovatively introduced the indicator of marital status, which has been proved to have an impact on the prognosis of patients with tumor in a number of studies.

⇒ A variety of factors, including race, gender, treatment style, pathological stage, and marital status were summarized, and their influence was comprehensively quantified.

How might these results change the focus of research or clinical practice?

⇒ All patients diagnosed with early stage GC can use our nomogram to assess the prognostic risk after receiving corresponding treatment.

⇒ Patients with high risk may receive relevant adjuvant therapy and moderately increase the frequency of physical examination.

⇒ In relevant policies, we should provide more social help and care to the widowed or single people.

INTRODUCTION

Gastric cancer (GC) is the fifth most common cancer and third leading cause of cancer-related deaths globally, with over 1 million new cases of GC and about 780 000 deaths in 2018.[1–3] In the past few decades, GC has been a main factor that has increased disability-adjusted life years globally, especially in areas with a GC high incidence, such as Japan, China, and other Asian regions.[4 5] GC is approximately twofold to threefold higher in men than in women and is uniformly rare in young people aged <50 years,[6] with increasing incidence rates after 50 years of age. Early stage GC is defined as GC limited in the lamina propria, mucosa, or submucosa, regardless of lymph node metastasis. Early stage GC has a greater chance of successfully getting removed through radical resection than advanced GC, consequently having a better prognosis than that of the latter.

Therefore, it is essential to diagnose and treat GC early to improve prognosis.

Even for patients with early stage GC who underwent systematic treatment, accurately predicting GC prognosis is difficult. Therefore, it is meaningful to establish a reliable predictive model in combination with post-treatment indicators. We obtained a large amount of clinical data regarding patients with early stage GC from the Surveillance, Epidemiology, and End Results (SEER) database to acquire a large sample size and great authenticity and incorporate various research indicators. According to current research, tumor size, stage_N, histology, age at diagnosis, tumor location, and other factors can affect survival.[2 7 8] However, these indicators have limited clinical developments, such as the wide application of endoscopic surgery (endoscopic mucosal resection and endoscopic submucosal dissection (ESD)). These organ-sparing therapies remain problematic for cancer cells and metastatic lymph node residues[9–11]; they also make fast recovery.[12 13] Therefore, early GC treatment may also be a prognostic factor for patients with cancer. Moreover, there are also other indicators that are related to the prognosis of patients with cancer. Marital status, which is associated with prostate, cervical and rectal cancer,[14–16] has emerged as an innovative risk factor in recent years. Some reports studied the impact of race on survival[17 18] ; these reports studied multiple races and were different from traditional studies that had only focused on one race. Chemotherapy has been proven to be effective for treating GC in long-term clinical applications. In recent years, neoadjuvant chemotherapy has become an important part of the treatment of advanced GC,[19] but the effect of chemotherapy on early stage GC remains controversial.

To the best of our knowledge, few studies have focused on the effect of these early stage GC indicators. Therefore, we performed a nomogram that can assess the impact of various indexes comprehensively to provide a basis for the prediction of the overall survival (OS) of patients with early stage GC.

MATERIALS AND METHODS

Data source and patient selection

In this study, we acquired data from the SEER database of patients with GC to evaluate the degree of the aforementioned factors. A nomogram, which is stable and visible, was used in our data analysis. A nomogram is based on multivariate analyses and integrates multiple predictive indicators; it can be used to diagnose diseases and predict their incidence or progression. We built a prediction nomogram based on independent accurate GC predictors. This gave us the ability to select an optional therapeutic regimen for individual patients. Research was restricted to tumors limited to the lamina propria, mucosa, and submucosa. Exclusion criteria in our study were as follows: (a) benign or stromal tumors; (b) distant metastasis or distant lymph node metastases; (c) second malignant primary indicator; (d) unknown chemotherapy; and (e) unknown survival time. Finally, 3647 cases were screened and included in this study as Figure 5 figure 1 showed. They were randomly divided into two groups—training and testing sets—based on a 3:1 ratio, respectively, meaning that 2719 people were in the training set and 928 people were in the testing set.

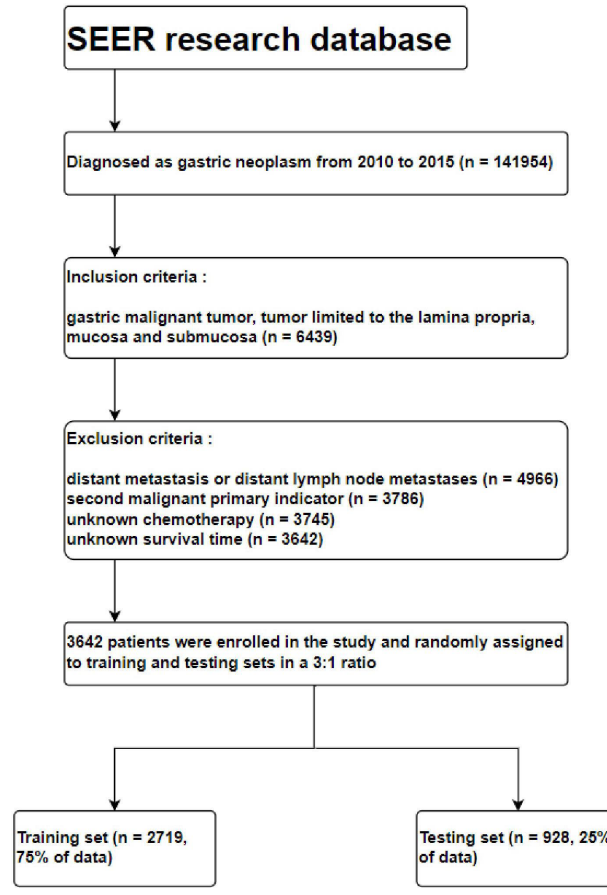

Figure 1

Data collection and end point

The following variables were included in our study: age at diagnosis, race, gender, tumor location, histology, grade, stage_T and stage_N, surgery in the primary site, lymph node dissection, chemotherapy, radiation, tumor size, insurance, and marital status. The main end point was OS, which was defined as the time from diagnosis until death due to any reason.

Age was divided into seven subgroups: ≤40, 40–50, 50–60, 60–70, 70–80, 80–90, and 90–100 years; race was divided into three subgroups: white, black, and other; the 'other' subgroup included Indian, Asian, and other minorities; tumor site was classified into eight subgroups according to the anatomy of the stomach: fundus, body, antrum, pylorus, lesser curvature, greater curvature, gastric overlapping area, and not otherwise specified (NOS); based on International Classification of Disease for Oncology, third edition, histology was divided into five subgroups: adenocarcinoma, signet ring cell carcinoma, special-type carcinoma, including carcinoid tumor, goblet cell carcinoid, and squamous cell carcinoma, other carcinomas, including neoplasms, diffuse type carcinoma, and linitis plastica, and unknown. Surgery was divided into five subgroups: no cancer-direct surgery, endoscopic surgery, partial gastrectomy, total gastrectomy, and unknown; lymph node dissection was divided into four subgroups: none, one to three regional lymph nodes removed, four or more regional lymph nodes removed, and unknown; and tumor size was

divided into eight subgroups: invisible to the naked eyes, ≤1 cm, 1–2 cm, 2–3 cm, 3–4 cm, 4–5 cm, >5 cm or widespread, and unknown. Marital status was divided into six subgroups: married (including domestic partner), divorced, separated, widowed, single (never married), and unknown.

Statistical methods

Continuous variables were expressed as mean±SD. Categorical variables were identified by frequency and proportion, which were both analyzed by Student's t-test and Pearson's χ^2 and Fisher's exact tests. A Kaplan-Meier analysis was performed to describe and compare survival among different variables, and parameters included mean and median survival times with a 95% CI. We also performed the log-rank test to compare the significance of survival curves. In the Cox proportional hazards regression analysis, variables that were considered significant in the univariate analysis were put in the multivariate analysis. These indicators that were ultimately meaningful were used to establish a nomogram to predict 3-year and 5-year OS. The parameters of the Cox proportional hazards regression analysis included HRs and 95% CIs. The C-index was employed to measure the reliability of the nomogram. We also built calibration curves to examine outcomes. All data were analyzed using SPSS (V.23.0) and R software (V.3.4.3).

RESULT

Baseline characteristics

A total of 141 954 patients were extracted from the SEER database, and 3647 suitable patients with early stage GC were included in this study. We divided patients into two cohorts—training (n=2719, 75% of data) and testing sets (n=928, 25% of data). Of the included patients, 1793 (49.2%) were male and 1854 (50.8%) were female. Moreover, 2231 were white, 607 were black, and 809 were put in the 'other' race subgroup. Regarding marital status, 1957 were married, 274 were divorced, 41 were separated, 630 were widowed, 512 were single, and 233 were classified into the 'unknown' group. Baseline characteristics of the training set is shown in online supplemental table 1. Age at diagnosis (p<0.001), race (p<0.001), gender (p<0.001), histology (p=0.007), grade (p=0.009), stage_T (p=0.025) and stage_N (p<0.001), surgery (p<0.001), tumor size (p=0.005), and insurance (p<0.001) were significantly different among marital status groups.

Kaplan-Meier survival analysis of marital status groups

To explore the influence of different marital status groups on OS, the Kaplan-Meier survival analysis was performed in all patients in the training set. As shown in figure 2, married individuals had the best prognosis (average OS=72.084, 95% CI=70.847 to 73.321), and the OS of widows was the worst (average OS=60.150, 95% CI=57.057 to 63.244). To verify whether gender is related to the above-mentioned results, we conducted the Kaplan-Meier survival analysis in patients with GC of different genders. As shown in figure 3A,B, there were significant differences in OS between sexes (p<0.001). In both male and female patients with early stage GC, survival was highest for married individuals (male average, OS=69.187, 95% CI=67.446 to 70.928; and female average, OS=76.357, 95% CI=74.783

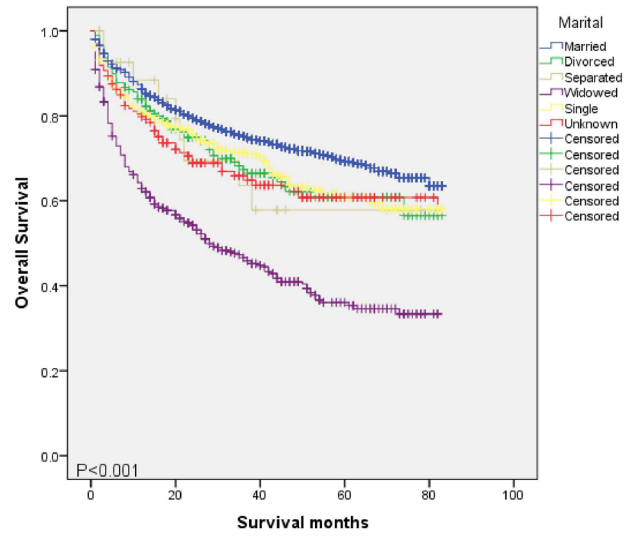

Figure 2 Kaplan-Meier survival analysis of overall survival among different marital status groups in patients with early stage gastric cancer (p<0.001).

to 77.930), and survival was the worst in widows (male average, OS=51.704, 95% CI=45.206 to 58.202; and female average OS=61.885, 95% CI=58.476 to 65.293). It is worthy to note that survival was significantly better in divorced female patients than in divorced male patients. Simultaneously, we also performed the Kaplan-Meier analysis of each known marital status group among genders, except for the 'separated' group as it had a small sample size and was consequently of limited reference. As shown in figure 4, there were significant differences between male and female patients in each marital status group (married, p<0.001, figure 4A; divorced, p=0.020, figure 4B; widowed, p=0.025, figure 4C; and single, p=0.026, figure 4D). Survival was better in female patients than in male patients.

PROGNOSTIC FACTORS OF PATIENTS WITH GC

Univariate analysis results are shown in table 1. The analysis showed that age at diagnosis, gender, histology, stage_T and stage_N, surgery, lymph node dissection, chemotherapy, radiation, tumor size, and marital status were significant prognostic factors. These univariate analysis factors were included in the multivariate analysis. Multivariate results showed that age at diagnosis, sex, histology, stage_T, surgery, tumor size, and marital status were independent prognostic factors for OS (table 2).

Prognostic nomogram for OS

Based on Cox regression models, a nomogram was constructed to predict the 3-year and 5-year OS of patients with early stage GC (figure 5). This nomogram created a scoring system in which each included variable can obtain a corresponding score of 0–100 according to their contribution to OS. After these scores were added to calculate the total score, the corresponding OS was predicted based on the scale at the bottom of the figure. This nomogram showed that tumor size was the most important prognosis factor, followed by age at diagnosis and surgery. Stage_T,

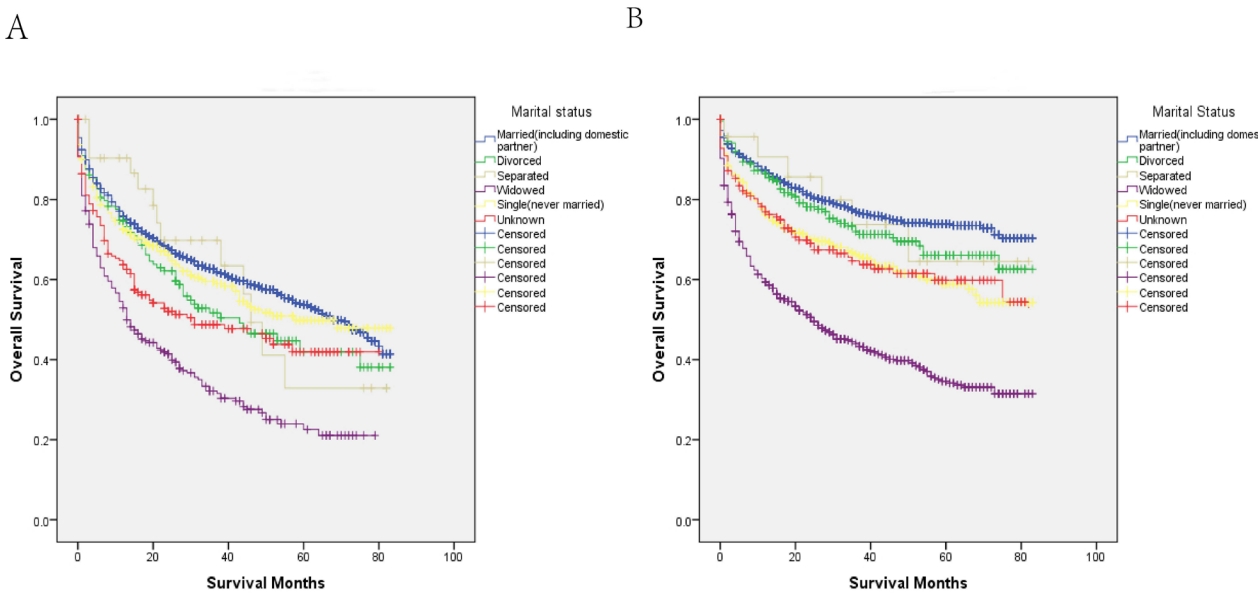

Figure 3 Kaplan-Meier survival analysis of overall survival among different marital status groups in genders. Overall survival among different marital status groups in (A) male patients (p<0.001) and (B) female patients (p<0.001) with early stage gastric cancer.

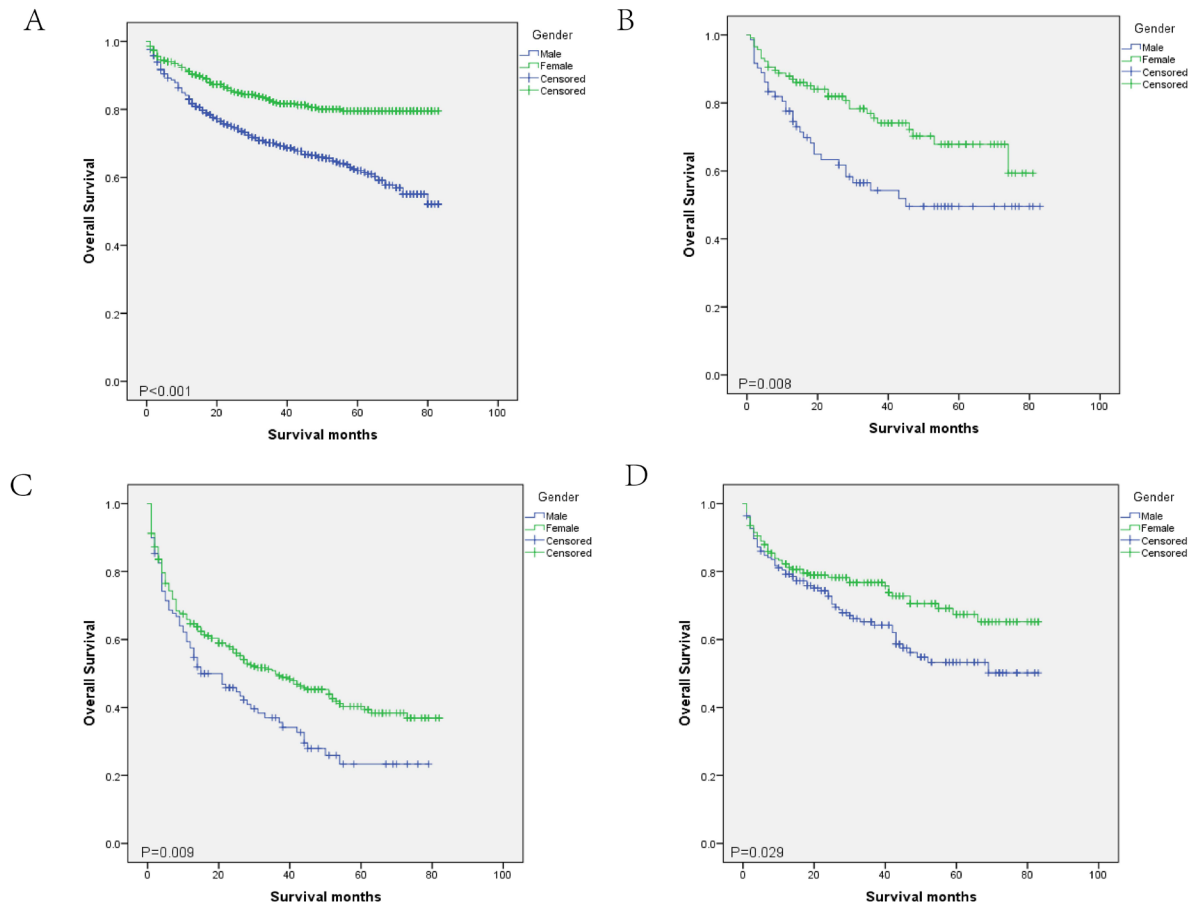

Figure 4 Kaplan-Meier survival analysis of each known marital status group among different genders, except for the 'separated' group. Overall survival (A) between married male and female patients (p<0.001), (B) between divorced male and female patients (p=0.008), (C) between widowed male and female patients (p=0.009), and (D) between single male and female patients (p=0.029).

Table 1 Univariate analysis of patients with early stage gastric cancer

Variables	HR (95% CI)	P value
Statistically significant factors		
Age at diagnosis (years)		<0.001
≤40 vs 41–50	0.048 (0.019 to 0.122)	<0.001
≤40 vs 51–60	0.086 (0.050 to 0.147)	<0.001
≤40 vs 61–70	0.110 (0.073 to 0.166)	<0.001
≤40 vs 71–80	0.133 (0.091 to 0.195)	<0.001
≤40 vs 81–90	0.213 (0.148 to 0.305)	<0.001
≤40 vs 91–100	0.462 (0.322 to 0.663)	<0.001
Race		0.097
White versus black	1.169 (0.938 to 1.456)	0.165
White versus other	1.363 (1.028 to 1.807)	0.031
Gender (male vs female)	0.690 (0.579 to 0.823)	<0.001
Location		0.488
Fundus versus body	1.260 (0.864 to 1.838)	0.230
Fundus versus antrum	1.023 (0.753 to 1.389)	0.886
Fundus versus pylorus	1.129 (0.862 to 1.479)	0.379
Fundus versus lesser curvature	1.018 (0.555 to 1.869)	0.953
Fundus versus greater curvature	1.107 (0.793 to 1.546)	0.551
Fundus versus overlapping	1.210 (0.816 to 1.794)	0.344
Fundus versus NOS	1.539 (1.045 to 2.266)	0.029
Histology		<0.001
Adenocarcinoma versus signet ring cell carcinoma	3.039 (2.054 to 4.496)	<0.001
Adenocarcinoma versus special type	2.122 (1.365 to 3.297)	0.001
Adenocarcinoma versus unknown	0.866 (0.515 to 1.457)	0.588
Adenocarcinoma versus other	2.893 (1.733 to 4.829)	<0.001
Grade		<0.001
Well differentiated versus moderately differentiated	0.618 (0.457 to 0.836)	0.002
Well differentiated versus undifferentiated	1.450 (1.139 to 1.846)	0.003
Well differentiated versus unknown	1.027 (0.807 to 1.306)	0.828
Stage_T		0.012
T1a versus T1b	0.731 (0.592 to 0.903)	0.004
T1a versus T1NOS	0.829 (0.673 to 1.020)	0.077
Stage_N		<0.001
N0 vs N1	0.550 (0.351 to 0.861)	0.009
N0 vs N2	0.962 (0.576 to 1.608)	0.884
N0 vs N3	0.870 (0.460 to 1.645)	0.668
N0 vs Nx	0.893 (0.378 to 2.112)	0.797
Surgery		<0.001
No cancer-direct surgery versus endoscopic surgery	1.554 (0.824 to 2.929)	0.173
No cancer-direct surgery versus partial gastrectomy	0.238 (0.121 to 0.469)	<0.001
No cancer-direct surgery versus total gastrectomy	0.295 (0.156 to 0.559)	<0.001
No cancer-direct surgery versus unknown	0.380 (0.196 to 0.739)	0.004
Lymph node dissection		<0.001
None vs 1–3 regional lymph nodes removed	1.321 (0.703 to 2.480)	0.387
None vs 4 or more regional lymph nodes removed	0.955 (0.456 to 1.996)	0.902
None versus unknown	0.653 (0.345 to 1.235)	0.190
Chemotherapy (none vs yes)	2.105 (1.718 to 2.581)	<0.001
Radiation (none vs yes)	2.172 (1.682 to 2.803)	<0.001
Size		<0.001
Invisible to the naked eyes vs ≤1 cm	0.049 (0.007 to 0.350)	0.003
Invisible to the naked eyes vs 1–2 cm	0.204 (0.158 to 0.264)	<0.001
Invisible to the naked eyes vs 2–3 cm	0.318 (0.244 to 0.415)	<0.001
Invisible to the naked eyes vs 3–4 cm	0.483 (0.358 to 0.652)	<0.001
Invisible to the naked eyes vs 4–5 cm	0.679 (0.490 to 0.939)	0.019
Invisible to the naked eyes vs >5 cm or widespread	0.623 (0.413 to 0.941)	0.024
Invisible to the naked eyes versus unknown	0.804 (0.570 to 1.134)	0.214
Insurance		0.626
Insured versus uninsured	0.799 (0.505 to 1.264)	0.338
Insured versus unknown	0.776 (0.377 to 1.599)	0.492
Marital		<0.001
Married versus divorced	0.811 (0.555 to 1.186)	0.280
Married versus separated	1.096 (0.687 to 1.747)	0.701
Married versus widowed	1.725 (0.843 to 3.530)	0.136
Married versus single	1.945 (1.306 to 2.896)	0.001
Married versus unknown	0.863 (0.557 to 1.338)	0.511

NOS, not otherwise specified.

marital status, gender, and histology also have a moderate impact on the prognosis of patients with early stage GC. The nomogram obtained in our study had good predictive ability and reliability.

Validation of the nomogram

In this study, we built a model that can predict the prognosis of patients with GC based on the SEER database; this model was validated by the testing set. The C-index was 0.791 and 0.685 in the training and testing sets, respectively, demonstrating that our nomogram was useful for patients with GC. Simultaneously, a calibration curve was used to examine the nomogram's ability to predict the 3-year and 5-year OS of patients of training and testing sets. As shown in online supplemental figure 1, the prediction of the nomogram was closely related to the observed results. We also performed a receiver operating characteristic curve; the 3-year survival area under the curve (AUC) was 0.774 and 0.717 in training and testing sets respectively, and the 5-year AUC was 0.773 and 0.722, as shown in online supplemental figure 2.

Kaplan-Meier curves for nomogram

According to scoring results, we divided patients into high-risk and low-risk groups (high-risk group and low-risk group were bounded by the median of the risk score) and performed the Kaplan-Meier survival analysis on these groups. As shown in online supplemental figure 3, there was a significant difference between the Kaplan-Meier curves of the high-risk group and those of the low-risk group, further demonstrating the reliability of the nomogram.

Table 2 Multivariate analysis of patients with early stage GC

Variables	HR (95% CI)	P value
Statistically significant factors		
Age at diagnosis (years)		<0.001
≤40 vs 41–50	0.048 (0.019–0.122)	<0.001
≤40 vs 51–60	0.086 (0.050–0.147)	<0.001
≤40 vs 61–70	0.110 (0.073–0.166)	<0.001
≤40 vs 71–80	0.133 (0.091–0.195)	<0.001
≤40 vs 81–90	0.213 (0.148–0.305)	<0.001
≤40 vs 91–100	0.462 (0.322–0.663)	0.303
Gender (male vs female)	0.690 (0.579–0.823)	<0.001
Histology		0.001
Adenocarcinoma versus signet ring cell carcinoma	3.039 (2.054–4.496)	0.001
Adenocarcinoma versus special type	2.122 (1.365–3.297)	0.001
Adenocarcinoma versus unknown	0.866 (0.515–1.457)	0.763
Adenocarcinoma versus other	2.893 (1.733–4.829)	0.068
Stage_T		<0.001
T1a versus T1b	0.731 (0.592–0.903)	<0.001
T1a versus T1NOS	0.829 (0.673–1.020)	0.045
Stage_N		0.052
N0 vs N1	0.550 (0.351–0.861)	0.483
N0 vs N2	0.962 (0.576–1.608)	0.419
N0 vs N3	0.870 (0.460–1.645)	0.338
N0 vs Nx	0.893 (0.378–2.112)	0.291
Surgery		<0.001
No cancer-direct surgery versus endoscopic surgery	1.554 (0.824–2.929)	0.472
No cancer-direct surgery versus partial gastrectomy	0.238 (0.121–0.469)	0.044
No cancer-direct surgery versus total gastrectomy	0.295 (0.156–0.559)	0.015
No cancer-direct surgery versus unknown	0.380 (0.196–0.739)	0.101
Lymph node dissection		0.083
None vs 1–3 regional lymph nodes removed	1.321 (0.703–2.480)	0.593
None vs 4 or more regional lymph nodes removed	0.955 (0.456–1.996)	0.276
None versus unknown	0.653 (0.345–1.235)	0.788
Chemotherapy (none vs yes)	2.105 (1.718–2.581)	0.132
Radiation (none vs yes)	2.172 (1.682–2.803)	0.597
Size		0.010
Invisible to the naked eyes vs ≤1 cm	0.049 (0.007–0.350)	0.025
Invisible to the naked eyes vs 1–2 cm	0.204 (0.158–0.264)	0.003
Invisible to the naked eyes vs 2–3 cm	0.318 (0.244–0.415)	0.018
Invisible to the naked eyes vs 3–4 cm	0.483 (0.358–0.652)	0.426
Invisible to the naked eyes vs 4–5 cm	0.679 (0.490–0.939)	0.917
Invisible to the naked eyes vs >5 cm or widespread	0.623 (0.413–0.941)	0.352
Invisible to the naked eyes versus unknown	0.804 (0.570–1.134)	0.682
Marital		<0.001
Married versus divorced	0.811 (0.555–1.186)	0.180
Married versus separated	1.096 (0.687–1.747)	0.018
Married versus widowed	1.725 (0.843–3.530)	0.001
Married versus single	1.945 (1.306–2.896)	<0.001
Married versus unknown	0.863 (0.557–1.338)	0.100

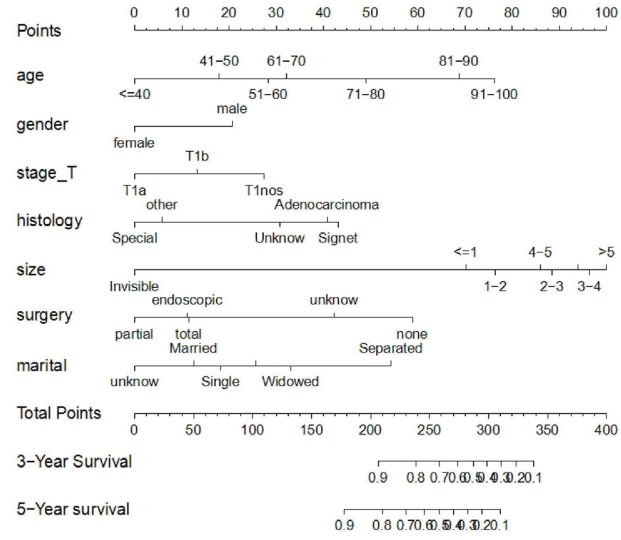

Figure 5 Nomogram predicting the overall survival of patients with early stage gastric cancer.

DISCUSSION

GC has two of the highest morbidity and fatality rates among cancers, originating from the gastric mucosal epithelium.[2] It can grow in various sites of the stomach and can easily develop hematogenous or lymphatic metastases.[3] In recent years, GC started to occur in young patients.[1] It is known that, even at early stages, GC may recur or develop metastases; therefore, it is important to maintain routine treatment and reviews to prolong patient survival.[20] But excessive treatment and examination will increase the financial burden on patients; however, it will affect GC prognosis. For example, enhanced CT, which is effective in diagnosing GC, is expensive and extremely unhealthy. Therefore, it is important to build a reliable nomogram that can accurately evaluate the recurrence risk of patients with GC postoperatively. Many studies revealed few GC prognostic factors, such as tumor size and invasion depth. However, these factors were limited and focused only on tumor growth and not on the patients' general condition and treatment information. Our research was based on the SEER database and included different races, innovatively adding some indicators that were proven to be associated with many kinds of cancer[14 16]; such indicators, such as marital status, are rarely used for GC. Although some studies have used nomograms to predict the prognosis of patients with GC,[21–23] we attempted to establish a prognostic nomogram combining multifarious clinical indicators, pathological characteristics, and treatment information to evaluate the probability of 3-year and 5-year OS of such patients.

In our study, the nomogram was more credible and persuasive as the outcomes were obtained from the data of the training set and then validated by testing set. First, we performed a univariate analysis including all factors; of these factors, we selected those that were significant, including age, sex, histology, and surgery, and brought them into the multivariate analysis. The multivariate analysis revealed that age at diagnosis, sex, histology, stage_T, surgery, tumor size, and marital status were independent prognostic factors of OS. A nomogram was constructed based on these factors,

and the C-index was 0.791. Calibration curves showed great consistency between prediction and observation results, and there was a significant difference between the high-risk and low-risk groups. Moreover, the AUCs of 3-year and 5-year survivals were 0.774 and 0.773, respectively.

The nomogram has been continuously proven to be a reliable and accurate prognostic prediction tool in recent studies. It can evaluate survival using various comprehensive indicators and acquire a better prediction effect than other prediction tools. For patients with early stage GC, based on the nomogram obtained in this study, combined with clinical information, we can obtain a postoperative patient risk rating. For high-risk patients, review frequency and follow-up times should be increased. Patients themselves should pay more attention to symptom fluctuation and improvements in lifestyle.

From the seven factors included in our nomogram, tumor size was the largest contributor to OS. This is in line with our usual perception, which is that a larger tumor is more aggressive and that a barely visible carcinoma in situ is indolent. Ohashi *et al* thought that both tumor size and depth could be used as combined prognostic indicators.[24] Our scoring system also included tumor invasion depth, and the T1b score was moderately higher than the T1a score. Tumors with a higher T stage have deeper infiltration, and there are more vascular and lymphatic vessels in the submucosa than in the mucosa, causing tumor cells to spread further and making them more difficult to remove; this directly worsens patient survival. Age was also an OS risk factor. Looking at the overall trend, old patients scored higher nomogram scores and had a worse prognosis than young patients. This might be attributed to the fact that elderly patients have a worse general condition and immune tumor cell clearance and more underlying diseases than the former.

It is worthy to note that different surgical methods also have a certain impact on prognosis. Patients who did not undergo surgery had the worst prognosis, indicating that surgery is still the most effective treatment for GC. Patients who underwent partial gastrectomy scored best on the nomogram, while patients who underwent endoscopic surgery and total gastrectomy had similar scores. This does not mean that total gastrectomy is not effective for treating GC because the condition of patients who need total gastrectomy might be more serious. Whether partial gastrectomy or endoscopic surgery is better for early stage small-diameter GC has remained a controversial issue in clinical practice, and a few studies have been dedicated to provide references to choose the correct treatment. Nishizawa and Yahagi indicated that patients receiving ESD generally had a better quality of life postoperatively; however, they also had a higher incidence of metachronous GC.[25] Mun *et al* reported that endoscopic surgery has fewer complications than traditional surgery based on the fact that the OS of the former is no less than that of the latter.[12] Nomura and Okajima suggested that we should try to reduce the extent of gastrectomy if GC can still be cured.[26] Our nomogram showed that partial gastrectomy was generally better than endoscopic surgery, indicating that, according to the current medical level, we should be cautious in using endoscopic surgery instead of traditional surgery. For patients with large tumor diameters and poor histological types, partial gastrectomy should be preferred to ensure radical resection. Because of its lower trauma rate and higher safety, endoscopic surgery can be applied to patients who cannot tolerate surgery due to advanced age or underlying diseases.

Marital status, a factor that is rarely enrolled in GC research, also showed moderate influence on survival in our nomogram. Married patients had the best prognosis, followed by single patients, and the prognosis of separated patients was the worst. This result was consistent with the Kaplan-Meier survival curve. We speculate that this might be due to the fact that married patients had better financial conditions and emotional encouragement, while separated patients might be more likely to experience financial difficulties emotional loss. Previous studies have shown that lower social support had a poorer prognosis for patients with cancer,[27] and marriage, one of the most social factor, was related to the prognosis of early GC in our study, this could be attributed to early diagnosis from the reminder and supervision of their partner. And our research showed that the married patients have the highest proportion of T1a and N0 than other group. Besides, research reported unmarried patients were less likely to have chemotherapy in patients with cancer (787/1360, 57.9%).[28] Therefore, the single patients and widowed patients need more attention and social help. Our study showed that the prognosis of female patients are better than male patients, this was consistent with other study,[29] this might be related to genetic differences between men and women. Li *et al* [30] reported that the expression of different core genes and differences in pathways were associated with the variation observed among patients with GC of different races and sexes.[28] Different lifestyles as a result of different sexes might also affect prognosis. Careful support may be required to improve the prognosis of male patients. Considering histology, signet ring cell carcinoma had the worst survival according to the nomogram. This result is also consistent with traditional clinical knowledge. Riihimäki *et al* demonstrated that signet ring adenocarcinomas had a higher probability of metastasis within the peritoneum, bone, and ovaries than adenocarcinomas.[31] The higher risk of metastasis may be the reason for the worse prognosis in patients with signet ring cell carcinoma.

To the best of our knowledge, this is the first SEER-based nomogram combining comprehensive clinical indicators to predict OS in patients with early stage GC. However, our research has some limitations. To improve the reliability of our study, we divided the screened SEER data into training and testing sets at a ratio of 3:1; however, the validation of the local medical center data was still missing. Second, retaining unclearly classified data or data displayed as 'unknown' enlarged the scope of application of the nomogram, and increased mutual interference between data to a certain extent; this affected the accuracy of the nomogram. Third, some well-known risk factors of GC, such as family history, alcohol and *Helicobacter pylori* (HP) infection, were not enrolled. These indicators were scarce in the SEER database as it was difficult to acquire them. For example, there is no clear standard to determine whether the patient has a history of drinking alcohol based on the amount of alcohol consumed, the frequency of drinking, and the time of abstinence. Moreover, HP infection examinations are not routinely performed in

many areas, making it difficult for the data to be applied in large databases.

CONCLUSION

In conclusion, our nomogram included age at diagnosis, sex, T stage, histology, tumor size, surgery, and marital status as risk factors effectively predicted the prognosis of early stage GC. This nomogram can help assess the prognosis and treatment of patients with GC.

Contributors LZ and BZ made the contribution to the main work equally, they designed the study. LZ provided the databases and prepared supplemental figures 1–3. PL, AX, WH, and ZW provided help and analyzed the data. BZ prepared figures 1–4 and wrote the manuscript. LZ acts as guarantor and be responsible for the overall content. All authors contributed to the article and approved the submitted version.

Funding This study was funded by Natural Science Foundation of Anhui Province, No. 2 108085QH337.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository.

ORCID iD

Aman Xu http://orcid.org/0000-0001-8142-4478

REFERENCES

1 Bray F, Ferlay J, Soerjomataram I, *et al*. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394–424.
2 Venerito M, Link A, Rokkas T, *et al*. Gastric cancer - clinical and epidemiological aspects. *Helicobacter* 2016;21:39–44.
3 Smyth EC, Nilsson M, Grabsch HI, *et al*. Gastric cancer. *Lancet* 2020;396:635–48.
4 Lin J-T. Screening of gastric cancer: who, when, and how. *Clin Gastroenterol Hepatol* 2014;12:135–8.
5 Kobayashi M, Sato Y, Terai S. Endoscopic surveillance of gastric cancers after Helicobacter pylori eradication. *World J Gastroenterol* 2015;21:10553–62.
6 Recio-Boiles A, Babiker HM. Gastric Cancer. In: *StatPearls [Internet]*. Treasure Island, FL: StatPearls Publishing, 2020.
7 Bilici A. Treatment options in patients with metastatic gastric cancer: current status and future perspectives. *World J Gastroenterol* 2014;20:3905–15.
8 Liang Y, Wu L, Liu L, *et al*. Impact of extranodal tumor deposits on prognosis and N stage in gastric cancer. *Surgery* 2019;166:305–13.
9 Park SE, Kim SH, Kim SG, *et al*. Local or extragastric recurrence after incomplete endoscopic submucosal dissection of early gastric cancer: risk factors and the role of CT. *Abdom Radiol* 2018;43:3250–9.
10 Liu Q, Ding L, Qiu X, *et al*. Updated evaluation of endoscopic submucosal dissection versus surgery for early gastric cancer: a systematic review and meta-analysis. *Int J Surg* 2020;73:28–41.
11 Venerito M, Link A, Rokkas T, *et al*. Gastric cancer - clinical and epidemiological aspects. *Helicobacter* 2016;21(Suppl 1):39–44.
12 Mun YG, Choi M-G, Lim C-H, *et al*. Factors affecting endoscopic curative resection of gastric cancer in the population-based screening era. *Clin Endosc* 2018;51:478–84.
13 Ono H, Yao K, Fujishiro M, *et al*. Guidelines for endoscopic submucosal dissection and endoscopic mucosal resection for early gastric cancer. *Dig Endosc* 2016;28:3–15.
14 Khan S, Nepple KG, Kibel AS, *et al*. The association of marital status and mortality among men with early-stage prostate cancer treated with radical prostatectomy: insight into post-prostatectomy survival strategies. *Cancer Causes Control* 2019;30:871–6.
15 Machida H, Eckhardt SE, Castaneda AV, *et al*. Single marital status and infectious mortality in women with cervical cancer in the United States. *Int J Gynecol Cancer* 2017;27:1737–46.
16 Diao J-D, Wu C-J, Cui H-X, *et al*. Nomogram predicting overall survival of rectal squamous cell carcinomas patients based on the SEER database: a population-based STROBE cohort study. *Medicine* 2019;98:e17916.
17 Ellis L, Canchola AJ, Spiegel D, *et al*. Racial and ethnic disparities in cancer survival: the contribution of tumor, sociodemographic, institutional, and neighborhood characteristics. *J Clin Oncol* 2018;36:25–33.
18 Saumoy M, Schneider Y, Shen N, *et al*. Cost effectiveness of gastric cancer screening according to race and ethnicity. *Gastroenterology* 2018;155:648–60.
19 Coccolini F, Nardi M, Montori G, *et al*. Neoadjuvant chemotherapy in advanced gastric and esophago-gastric cancer. Meta-analysis of randomized trials. *Int J Surg* 2018;51:120–7.
20 Findlay JM, Antonowicz S, Segaran A, *et al*. Routinely staging gastric cancer with ^{18}F-FDG PET-CT detects additional metastases and predicts early recurrence and death after surgery. *Eur Radiol* 2019;29:2490–8.
21 Yu C, Zhang Y. Development and validation of prognostic nomogram for young patients with gastric cancer. *Ann Transl Med* 2019;7:641.
22 Bando E, Ji X, Kattan MW, *et al*. Development and validation of a pretreatment nomogram to predict overall survival in gastric cancer. *Cancer Med* 2020;9:5708–18.
23 He Y, Mao M, Shi W, *et al*. Development and validation of a prognostic nomogram in gastric cancer with hepatitis B virus infection. *J Transl Med* 2019;17:98.
24 Ohashi T, Komatsu S, Ichikawa D, *et al*. Tumor index as a combined indicator of tumor depth and size in gastric cancer. *Anticancer Res* 2016;36:1895–900.
25 Nishizawa T, Yahagi N. Long-term outcomes of using endoscopic submucosal dissection to treat early gastric cancer. *Gut Liver* 2018;12:119–24.
26 Nomura E, Okajima K. Function-preserving gastrectomy for gastric cancer in Japan. *World J Gastroenterol* 2016;22:5888–95.
27 Thompson T, Rodebaugh TL, Pérez M, *et al*. Influence of neighborhood-level factors on social support in early-stage breast cancer patients and controls. *Soc Sci Med* 2016;156:55–63.
28 Klapheke A, Yap SA, Pan K, *et al*. Sociodemographic disparities in chemotherapy treatment and impact on survival among patients with metastatic bladder cancer. *Urol Oncol* 2018;36:308.e19–308.e25.
29 Kiuchi J, Komatsu S, Ichikawa D, *et al*. [Differences in Related Risk Factors for Severe Postoperative Complications and Prognosis between Male and Female Gastric Cancer Patients]. *Gan To Kagaku Ryoho* 2016;43:1957–9.
30 Li H, Wang C, Wei Z, *et al*. Differences in the prognosis of gastric cancer patients of different sexes and races and the molecular mechanisms involved. *Int J Oncol* 2019;55:1049–68.
31 Riihimäki M, Hemminki A, Sundquist K, *et al*. Metastatic spread in patients with gastric cancer. *Oncotarget* 2016;7:52307–16.