BMC Cancer

RESEARCH

Open Access

Effect of comorbidity classes on survival of patients with gastrointestinal tract cancer

Linna Gao[1†], Tian Yao[2†], Shaohua Ge[3†], Jiuwei Cui[4], Wei Li[4], Zengqing Guo[5], Hongxia Xu[6], Min Weng[7], Suyi Li[8], Qinghua Yao[9], Wen Hu[10], Lan Zhou[11], Junqiang Chen[12], Xianghua Wu[12], Qingchuan Zhao[13], Hongli Li[14], Hanping Shi[15*], Yi Ba[16*], He Huang[1,17*] and The Investigation on Nutrition Status and its Clinical Outcome of Common Cancers (INSCOC) Group[1*]

Abstract

Background Comorbidities may complicate medical situations and have an impact on the treatment decisions and poor survival of cancer patients. How comorbidities cluster together and ultimately affect patients' outcomes in gastrointestinal tract cancer (GTC) is a poorly understood area.

Methods In a multicenter prospective observational study from 2012 to 2021, we grouped the comorbidities of patients with GTC by latent class analysis, obtaining two comorbidity classes. Cox regression models were initially used to predict mortality. LASSO techniques were used to reduce the dimension. The final model included the comorbidity classes and nine more predictors. Additionally, the performance of different simple multimorbidity measures were compared using the Bayesian information criterion (BIC), ROC curves and C-index. Finally, the performance of the final model was analyzed using ROC curves, calibration curves and decision curves. The nomogram was drawn to evaluate the model.

Results We included 10,019 patients and obtained two comorbidity classes. Class 2 patients have a higher incidence of comorbidities, and a lower survival rate compared to Class 1 ($P < 0.001$). Compared to models containing the number of comorbidities or only a single comorbidity, the final model with the comorbidity classes has the highest AUC and C-index, as well as the lowest BIC, indicating this model has the best predictive performance.

Conclusion We identified two classes of comorbidities that were associated with overall survival in patients with GTC. The combination of different comorbidities class plays a vital role in the prognosis of GTC.

Keywords Gastrointestinal tract cancer, Comorbidity, Predictive model, Overall survival

Background

Malignant gastrointestinal tract cancer (GTC) are significant diseases that seriously threaten human health. Esophageal cancer (EC), gastric cancer (GC) and colorectal cancer (CRC) ranked 10th, 6th, and 5th in the global cancer incidence in 2020, with the total number of new cases and deaths accounting for 29% and 33% of the top 10 malignancies [1]. The overall prognosis of patients with these cancers has improved in recent decades due to surgical and medical progress, but overall survival remains poor especially in EC and GC. The 5-year

† Linna Gao, Tian Yao and Shaohua Ge are co-first authors and contributed equally to this work.

*Correspondence:
Hanping Shi
shihp@ccmu.edu.cn
Yi Ba
yiba99@yahoo.com
He Huang
hh93003@163.com
The Investigation on Nutrition Status and its Clinical Outcome of Common Cancers (INSCOC) Group
Full list of author information is available at the end of the article

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

survival rates for EC, GC, CRC were 20% [2], 30% [3] and 65% [2].

The prognostic factors for GTC have been extensively reported. A large number of studies have focused on the impact of a single comorbidity [4] or the numbers of comorbidities [5–7]. It shows that comorbidities are important prognosis factors to gastrointestinal cancer patients [8]. A retrospective study found that almost half of colon cancer patients have three or more comorbidities [8], and those with a higher comorbidity burden are less likely to receive or complete standard anticancer treatments and have shorter overall survival [9–14]. However, the effect of comorbidity may be underestimated if we use simple measures. The pathological mechanisms of different comorbidities will have different effects on the prognosis. Accurately assessing the impact of comorbidities on prognosis should be based on a deep understanding of the types, the numbers, the combinations and interactions of different comorbidities and cancers. Therefore, the comprehensive assessment of comorbidities should be considered in clinical treatment decisions and patients' quality of life.

The objectives of this study were to identify distinct comorbidity classes and their effect on overall survival in patients from a multi-center, large-sample study, based on the Investigation on Nutrition Status and its Clinical Outcome of Common Cancers (INSCOC) data [15]. We then constructed a clinical prognostic model including comorbidity classes and validated its predictive performance in patients' survivals.

Materials and methods

Study population

Data were collected from the INSCOC project, spanning the period from January 2012 to May 2021. This project is a multi-center, large-sample, and prospective study conducted in China. The study protocol has been registered with the Chinese Clinical Trial Registry (ChiCTR), with the identifier ChiCTR1800020329 [16]. The inclusion criteria and exclusion criteria were seen in Extension protocol for the INSCOC study. This study ultimately included 10,019 patients from 45 hospitals in 20 provinces of China (Supplementary Table 1). The study design was approved by the local ethics committees of each participating hospital, and the project was conducted in accordance with the updated Declaration of Helsinki. Before participating in the study, all recruited patients signed a written agreement form authorizing the use of their data.

Data acquisition

The baseline information for all cancer patients was collected by one or two well-trained professional researchers at each institution. The information included demographic data, clinical data (cancer type, histological type, cancer stage, number of metastatic organs), laboratory data (including four composite inflammatory markers), Patient interview data (nutrition risk screening 2002 (NRS2002), Karnofsky Performance Status (KPS)), and treatment data (surgery, radiotherapy, chemotherapy, nutrition support). The response variable were the survival status and time of GTC patients.

Follow-up and quality control

Follow-up was performed by each center at the 6th month after enrolling, as well as every year thereafter. All follow-up were carried out by professionally trained staff. The final follow-up was conducted in June 2021. The primary endpoint was the all-cause mortality from enrollment, confirmed by the patients' family, hospital medical records, and healthcare professionals.

Follow-up and data quality were regularly checked and assessed at multiple centers as part of quality control procedure. Quality control included timely reporting of follow-up progress and regular assessment of follow-up staff. Data quality control included checking for data entry errors, omissions, and outliers, and promptly correcting corresponding any errors.

Statistical analysis

Latent Class Analysis (LCA) is a person-centered analytical approach that partitions a population into subgroups or classes based on a set of observed variables [17]. In this study, we utilized LCA to identify distinct comorbidity classes among patients (Supplementary file 1). The process began by specifying different numbers of classes—ranging from two to a maximum number (designated by researchers as an integer based on the data and research questions). To determine the optimal number of classes, we relied on several model selection criteria. The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were used, as they penalize model complexity to varying degrees, with lower values suggesting a better balance between model fit and complexity, which helped in avoiding overfitting by preferring simpler models [18]. The Lo-Mendell-Rubin adjusted likelihood ratio test (LMR) was used to compare models with successive numbers of classes. For example, if researchers set the number of classes to four, the LMR assesses whether a model with four classes significantly improves the model fit over a model with three classes. Entropy, a measure of classification precision, was also calculated for each model. High entropy values — ideally above 0.8 — indicate that the classification is reliable, with over 90% of patients being correctly classified into their respective comorbidity classes [19]. Relative Frequency (RF) denotes the

proportion of individuals within each latent class relative to the total sample. When the RF of a smallest class falls below 5%, it may suggest that the class is underrepresented within the sample, prompting a reevaluation of the model's rationality [18]. These methods enable the translation of complex comorbidity profiles into comorbidity classes.

The participants were randomly assigned to the training and validation sets. Initially, the Least Absolute Shrinkage and Selection Operator (LASSO) method with 10-fold crossover was used to identify potential predictive factors and construct a final model, a method known for its ability to shrink less important variable coefficients to zero, effectively performing variable selection and enhancing the model's interpretability. Cox univariate regression analysis was conducted on all variables, and Cox multivariate regression analysis was conducted on the predictive factors selected by LASSO. Secondly, the Comorbidity class variable in the final model was replaced by the number of comorbidities or a single comorbidity to construct four new models (the three most serious common comorbidities that might have a substantive effect of survival were selected for analysis: coronary heart disease (CHD), COPD and diabetes). The methods of model comparison included the use of area under the curve (AUC) of the ROC curve (Receiver Operating Characteristic), BIC and concordance index (C-index). Thirdly, in the final model, a nomogram was constructed in order to predict 1-, 3- and 5-year overall survival, decision curve was generated in order to assess the net benefit of the model, the calibration curve was constructed to evaluate the accuracy of the probabilistic output of the prediction model.

Missing data represent less than 5% of the observations in the full dataset and were imputed using multiple interpolations. For all analysis, $p \leq 0.05$ was considered statistically significant. Categorical data were presented as frequencies and proportions. Continuous variable is described using the median, along with the 25th percentile and 75th percentile. The survival time of patients is described using the mean ± standard deviation (SD). Chi-square test and Fisher exact test were used to compare the rates between the two groups. The Kaplan-Meier (K-M) survival curve was used to describe the survival of patients in each group. LCA models were estimated in Mplus Version 8.3 (Muthén & Muthén, 2020) [20]. Other statistical analysis performed were conducted using SPSS 29.0.1 (SPSS Software, Chicago, USA) [21] and R 4.3.2 (R Foundation for Statistical Computing, Vienna, Austria) [22]. The R package used in this study includes glmnet [23], rms [24], VIM [25], survival [26], dcurves [27], autoReg [28], ggplot2 [29], timeROC [30].

Results

Baseline characteristics

A total of 10,019 hospitalized GTC patients (EC $n=2156$; GC $n=3057$; CRC $n=4806$) were included in the analysis (Supplementary Table 2). The average follow-up of patients with esophageal cancer, gastric cancer and colorectal cancer was 31.3 ± 23.7 months, 38.4 ± 24.4 months and 39.0 ± 24.3 months, respectively. The average follow-up time of the total population was 37.2 ± 24.4 months (median 35 months), the longest follow-up was 97 months, and the most times of follow-up visits was seven. There were 2950 (29.4%) patients died during follow-up. The median age at diagnosis was 59 years, and 52.6% were diagnosed at < 60 years; 68.1% were male, and 48.8% underwent surgery. The 1-, 3-, 5-year survival rates of the overall population were 87.4%, 74.7%, and 65.7%.

Comorbidity classes

All models had Entropy > 0.8 and LMR $p < 0.001$, indicating that the classification accuracy of the models was greater than 90%. Then, we assessed the relative RF of the smallest class for each model and discovered that it was too small in 3-class model (RF = 0.0096) and 4-class model (RF = 0.0026), as detailed in Table 1. It meant that the population was too small to be representative, so we chose the 2-class model.

The median age was not different between the two classes. Class 1 included relatively healthy patients or with fewer comorbidities, accounting for 91.3% of the total population. Among Class 1 patients, 75.2% had no comorbidities, 23.1% had only one comorbidity, and 1.7% had two comorbidities. None of the patients in this group had stroke, myocardial infarction, chronic pancreatitis, osteoporosis, or renal disease. Class 2, although accounting for 8.7% of the total population (870 individuals), was mainly characterized by having multiple comorbidities (Supplementary Table 3 and Supplementary Table 4). Among Class 2 patients, 3.0% had only one comorbidity, 53.7% had two comorbidities, and 22.6% had three comorbidities, and 20.7% had four or more comorbidities. In addition, except for

Table 1 Latent class Model Fit comparison ($N=10019$)

	AIC	BIC	Entropy	LMR *p* value	RF
1-class model	34,370	34,471	-	-	-
2-class model	32,120	32,329	0.811	< 0.001	0.0868
3-class model	31,866	32,183	0.842	< 0.001	0.0096
4-class model	31,707	32,133	0.869	< 0.001	0.0026

AIC Akaike information criterion, *BIC* Bayesian information criterion, *LMR* Lo-Mendell-Rubin adjusted likelihood ratio test, *RF* Relative frequency for smallest class

Tuberculosis and Hepatic disease, which were not discriminatory between the two categories, the incidence of comorbidities was much higher in Class 2 than in Class 1. Of these, the incidence of inflammatory bowel disease in Class 2 was 5.2 times higher than that in Class 1, COPD was 8.6 times, diabetes was 10.9 times, and coronary heart disease (CHD) was 23.8 times. KM survival curves showed that the survival of patients in Class 2 was significantly shorter than those in Class 1 ($P < 0.001$) (Supplementary Fig. 1). The 1-, 3-, 5-year survival rates in Class 1 were 87.6%, 75.0%, 66.6%, and in Class 2 were 85.1%, 71.3%, 55.7%.

To explore the potential differences between comorbidity classes, we conducted univariate analyses targeting other factors (Supplementary Table 5). We found that Class 1 had a significantly higher proportion of males, smokers, and alcohol drinkers, as well as patients with only a compulsory education, with a normal weight, without a family history, and residing in rural areas, compared to Class 2 (all $P < 0.001$). Class

Table 2 Univariate and multivariate analysis

Variable	Value	Univariate		multivariate	
		HR (95% CI)	*P*	HR (95% CI)	*P*
Sex	Female	0.824 (0.749, 0.907)	< 0.001		
Age	50–60	1.206 (1.056, 1.377)	0.006	1.168 (1.020, 1.336)	0.024
Age	60–70	1.411 (1.242, 1.603)	< 0.001	1.302 (1.142, 1.484)	< 0.001
Age	≥ 70	1.917 (1.665, 2.207)	< 0.001	1.563 (1.349, 1.810)	< 0.001
Smoking	Ever	1.235 (1.132, 1.347)	< 0.001	1.140 (1.042, 1.248)	0.004
Drinking	Ever	1.098 (0.996, 1.212)	0.060		
FH	Yes	1.001 (0.884, 1.133)	0.988		
Place	Rural areas	1.023 (0.938, 1.116)	0.604		
Education	Above compulsory education	0.878 (0.800, 0.963)	0.006		
Occupation	Physical work	0.899 (0.808, 1.001)	0.052		
Occupation	Mental work	0.774 (0.677, 0.883)	< 0.001		
Occupation	Other work	1.005 (0.882, 1.146)	0.936		
BMI	18.5–24	0.680 (0.611, 0.757)	< 0.001		
BMI	≥ 24	0.618 (0.542, 0.705)	< 0.001		
NRS2002		1.172 (1.142, 1.202)	< 0.001	1.111 (1.081, 1.141)	< 0.001
KPS		0.987 (0.985, 0.990)	< 0.001	0.990 (0.987, 0.993)	< 0.001
NLR		1.001 (0.999, 1.004)	0.234		
PLR		1.000 (1.000, 1.000)	0.508		
SII		1.000 (1.000, 1.000)	0.046		
PNI		0.997 (0.995, 1.000)	0.020		
Hospital stays		1.002 (0.999, 1.006)	0.190		
Stage	Stage III + Stage IV	2.030 (1.798, 2.292)	< 0.001	1.590 (1.404, 1.801)	< 0.001
Cancer type	GC	0.761 (0.680, 0.850)	< 0.001	0.582 (0.448, 0.756)	< 0.001
Cancer type	CRC	0.535 (0.479, 0.596)	< 0.001	0.432 (0.331, 0.563)	< 0.001
Histological type	Squamous cell carcinoma	1.546 (1.401, 1.708)	< 0.001	0.752 (0.580, 0.974)	0.031
Histological type	Signet-ring cell carcinoma	1.756 (1.412, 2.183)	< 0.001	1.720 (1.374, 2.153)	< 0.001
MON	1	2.149 (1.920, 2.406)	< 0.001	2.074 (1.849, 2.327)	< 0.001
MON	≥ 2	3.194 (2.834, 3.599)	< 0.001	2.956 (2.615, 3.342)	< 0.001
Surgery	Yes	0.766 (0.702, 0.836)	< 0.001	0.885 (0.808, 0.968)	0.008
Radiotherapy	Yes	1.124 (0.978, 1.290)	0.099		
Chemotherapy	Yes	1.298 (1.190, 1.416)	< 0.001		
Nutrition support	Yes	0.986 (0.897, 1.084)	0.770		
Comorbidity class	Class 2	1.376 (1.200, 1.578)	< 0.001	1.324 (1.151, 1.523)	< 0.001

NRS 2002 nutrition risk screening 2002; *KPS* Karnofsky Performance Scale; *NLR* neutrophil-to-lymphocyte ratio; *PNI* the prognostic nutritional index; *MON* Metastatic organ number; *EC* Esophageal cancer; *GC* Gastric cancer; *CRC* Colorectal cancer; *AC* adenocarcinoma; *SCC* Squamous cell carcinoma; *SRCC* Signet-ring cell carcinoma

In the multivariate analysis, age, smoking status, NRS2002, KPS, cancer stage, cancer type, histological type, metastatic organ number, surgery status, and comorbidity class were found to be significant predictors for patient outcomes. The adjusted hazard ratios (HR) and their 95% confidence intervals (CI) are presented, with *P*-values indicating the level of statistical significance. A *P*-value of less than 0.05 was considered statistically significant

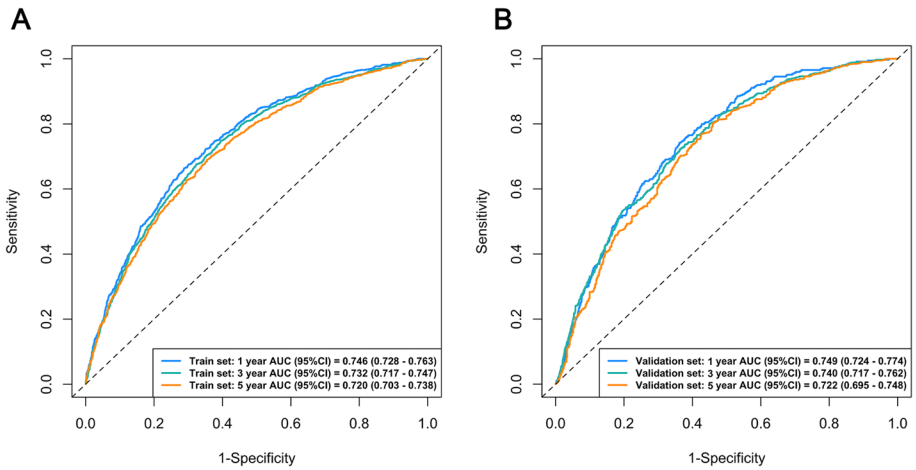

Fig. 1 ROC for training set (**A**) and validation set (**B**)

1 tended to include younger patients, with more than half of the patients being under 60 years old, and only 13.9% were 70 years old or older. In contrast, Class 2 had 33.3% of patients under 60 years old, and a notably higher proportion of patients aged 70 years and older, accounting for 27.9% of the class. And the proportion of retired patients in Class 2 (45.7%) was significantly higher than that in Class 1 (25.6%), which could be explained by the higher proportion of older patients in Class 2. In addition, we found no significant difference in cancer staging between Class 1 and Class 2. And there were more patients with colorectal cancer in Class 2 ($P=0.002$).

Training set and validation set

The proportions of training and validation sets were 70.0% ($n=7013$) and 30.0% ($n=3006$) in the patients. The mean follow-up for the training and validation sets were 37.4 ± 24.3 months and 36.5 ± 24.4 months. There were no significant differences in most baseline characteristics between two cohorts ($P>0.05$), suggesting an equilibrium distribution for the two cohorts (Supplementary Table 2). The KM survival curves showed no significant difference in overall survival between the training and validation sets ($P=0.19$) (Supplementary Fig. 2), indicating the even distribution between two sets and suitable for internal verification. In the training set, the 1-, 3- and 5-year survival rates for Class 1 were 87.7%, 75.4% and 67.1%, and for Class 2 were 85.3%, 71.4% and 55.3%. In the validation set, the 1-, 3- and 5-year survival rates for Class 1 were 87.4%, 74.1% and 65.4%, and for Class 2 were 84.6%, 71.2% and 56.8%.

Variable selection

The Lasso method (Supplementary Fig. 3 and Supplementary Fig. 4) was initially employed to reduce the number of candidate variables, resulting in the selection

Table 3 Comparison of different comorbidity measures

	Comorbidity classes model	Comorbidity count model	CHD model	COPD model	Diabetes model
Train set					
AUC-1year	0.746 (0.728–0.763)	0.622 (0.602–0.643)	0.621 (0.601–0.642)	0.621 (0.601–0.642)	0.624 (0.604–0.644)
AUC-3year	0.732 (0.717–0.747)	0.601 (0.584–0.618)	0.598 (0.580–0.615)	0.598 (0.580–0.615)	0.601 (0.584–0.618)
AUC-5year	0.720 (0.703–0.738)	0.628 (0.609–0.646)	0.619 (0.600–0.638)	0.616 (0.597–0.635)	0.622 (0.603–0.641)
BIC	33,359	33,935	33,936	33,943	33,928
C-index	0.704	0.601	0.598	0.597	0.599
Validation set					
AUC-1year	0.749 (0.724–0.774)	0.630 (0.599–0.661)	0.622 (0.591–0.654)	0.622 (0.591–0.654)	0.629 (0.597–0.660)
AUC-3year	0.740 (0.717–0.762)	0.614 (0.589–0.640)	0.603 (0.578–0.629)	0.606 (0.580–0.632)	0.612 (0.586–0.638)
AUC-5year	0.722 (0.695–0.748)	0.626 (0.596–0.655)	0.619 (0.589–0.648)	0.620 (0.590–0.649)	0.621 (0.592–0.650)

AUC area under the curve; *BIC* bayesian information criterion; *C-index* concordance index; *ROC* Receiver Operating Characteristic

We have renamed the final model with the comorbidity classes as the 'Comorbidity classes model' in this table to better distinguish it from the other four models

of 10 from the original 25, including age, smoking, NRS2002, KPS, stage, cancer type, histological type, the number of metastatic organs, surgery, and comorbidity classes. The results of the Cox regression analysis of these factors, as illustrated in Table 2, indicated that comorbidity class was an independent risk factor for predicting patient outcomes.

Comparison of comorbidities measures

Whether in the training or validation set, the final model has the highest AUC of the ROC curve, (Fig. 1), the lowest BIC, and the highest C-index (Table 3). These results suggest that the comorbidity classes offer a better performance than the basic measures.

Charasteristics of final model with the comorbidity classes

The nomogram (Fig. 2) provides an integrated perspective to evaluate the impact of different factors on the survival of cancer patients, helping to guide clinical decision making and patient management. The points on the calibration plot generally fall along the 45-degree line, indicating that the predicted probability of survival closely matches the actual probability of survival, which suggests that the model predicts correctly (Fig. 3. A-C). The decision curve (Fig. 3. D-F) showed different net benefits of the model at different threshold probabilities at 1, 3 and 5 years, which helped clinicians to make decisions accordingly.

Discussion

Of the 4.8 million new cases of gastrointestinal cancers and 3.4 million related deaths worldwide in 2018, it is estimated that 38% of the cases and 41% of the deaths occurred in China [31]. Complications are important factors affecting the prognosis of patients. Numerous studies have confirmed the effect of a single comorbidity on gastrointestinal cancers [32–35]. However, it is a critically understudied area to determine how comorbidities cluster together and affect prognosis.

In this study, we first used the LCA method and its parameters (AIC, BIC, Entropy, LMR p value, RF) to group the comorbidities of EC, GC, and CRC patients and identified two classes of comorbidities. Among them, Class 2 patients have a high incidence of comorbidities and a short survival period. Secondly, we compared the models containing the Comorbidity class with those containing only a single comorbidity or the number of comorbidities, demonstrating that the combination of comorbidities has a significant impact on the prognosis of gastrointestinal cancer patients. Therefore, our study

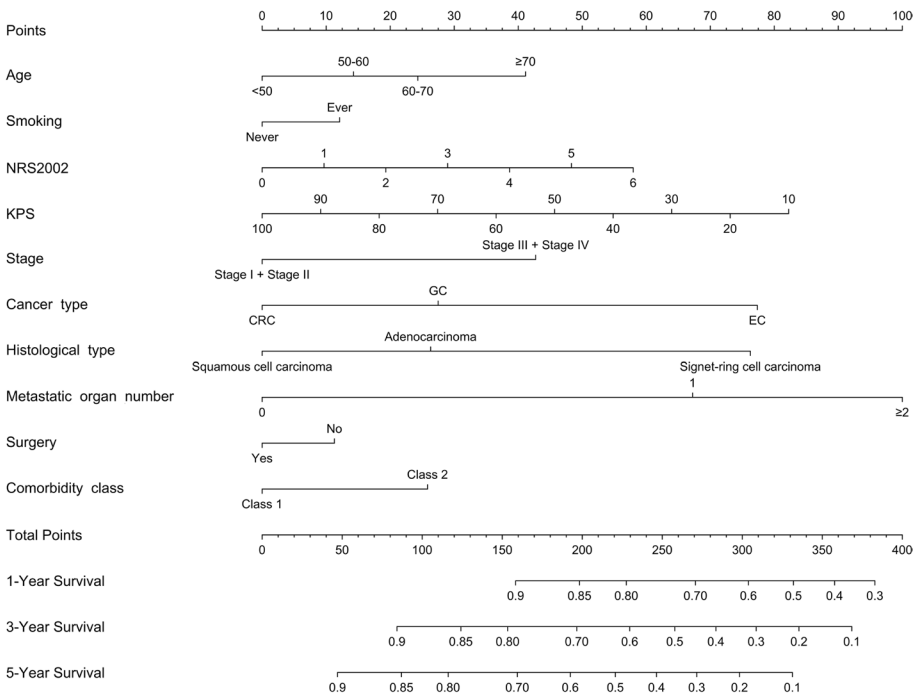

Fig. 2 Nomogram based on final model with the comorbidity classes Note: The steps on how to read and calculate the total integral of a nomogram: 1) Determine the specific value of each prognostic factor 2) Draw a vertical line for each prognostic factor, find the corresponding score on the Points line at the top of the Fig. 3) Add all corresponding scores to obtain the total score 4) The Total Points line at the bottom of the nomogram represents the relationship between total points and survival probability. Map the total score to this line, and obtain predicted survival probabilities for time points such as 1 year, 3 years, and 5 years

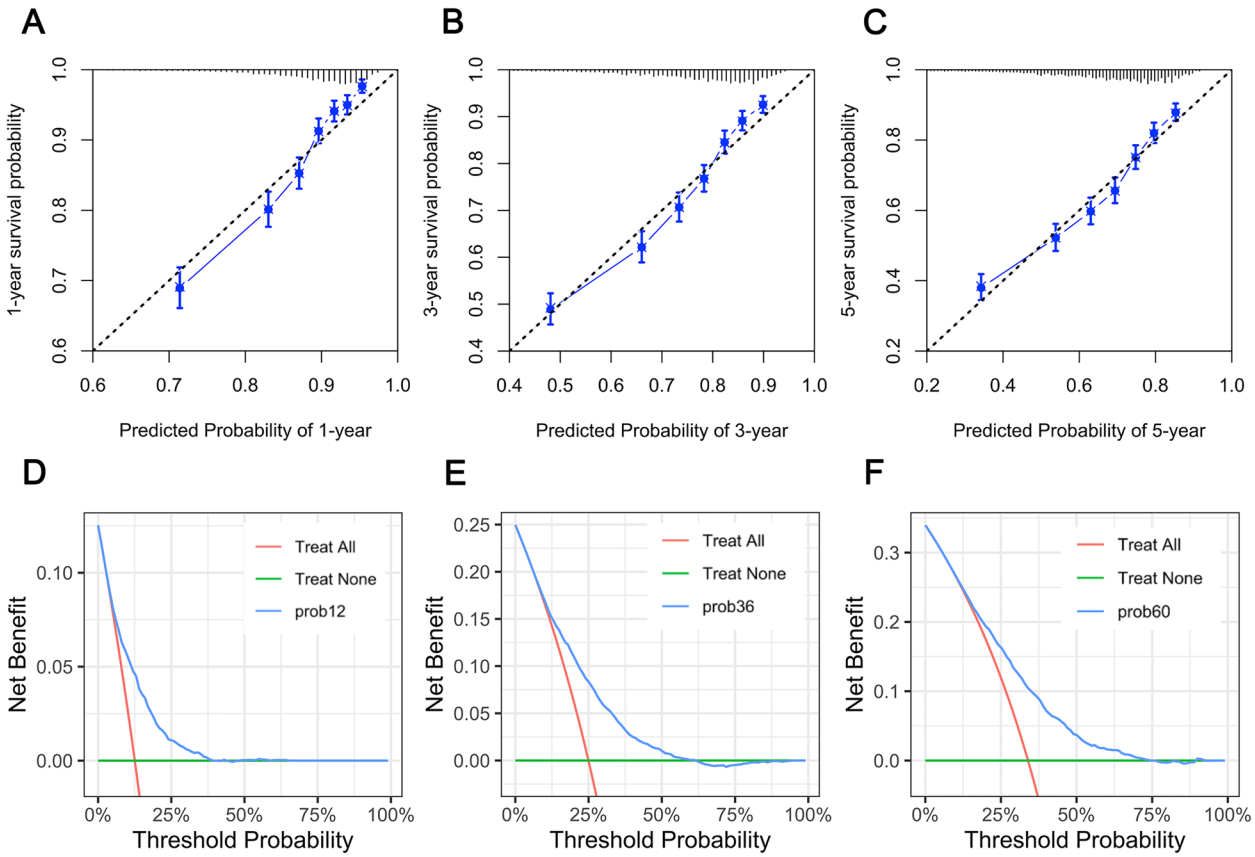

Fig. 3 Calibration curves for 1-year(A), 3-year(B), 5-year(C); Decision curves for 1-year(D), 3-year(E), 5-year(F)

emphasizes that when evaluating the impact of comorbidities on patient prognosis, individual comorbidities should not be examined in isolation, but their combined effects on the overall health status and survival of patients should be comprehensively considered.

Among the various comorbidities we collected, COPD, diabetes and CHD are dominant. Numerous findings have demonstrated an increased risk of death due to COPD in cancer patients and highlighted the importance of respiratory care throughout survival [36–38]. However, the relationship between cancer and COPD mortality remains unclear [39]. Studies have reported that risks of death and progression were greater for patients with diabetes than those without diabetes in gastrointestinal cancer patients [33, 40]. A consensus report from the American Diabetes Association and American Cancer Society points that the association between diabetes and cancer is not well established. Although a number of potential mechanisms have been proposed, such as insulin resistance and hyperglycaemia cancer promotion. The biological mechanisms by which diabetes increases cancer risk remain undefined [34]. The impact of CHD on the prognosis of patients with gastrointestinal cancers may come from obesity [41]. Obesity is a significant risk factor for CHD. The major systematic changes associated with obesity include the alteration of insulin, leptin, cytokines, and steroid hormones. These factors alter the nutritional environment, which may create conditions that trigger and sustain tumor growth [42].

The other predictive factors included in our model have been supported by relevant literature in their prognostic importance. In the overall population, signet ring cell carcinoma (SRCC) had the highest risk, which is consistent with previous studies. Signet ring cell carcinoma is a histological type based on microscopic characteristics that occurs in gastrointestinal cancer. In early-stage cancer, the prognosis of SRCC is similar [32, 35] or better than other adenocarcinoma [43, 44]. Among advanced gastric cancer patients, SRC patients had a worse prognosis than other cell types [45–47]. In our study, SRCC was an independent predictor of poor prognosis, consistent with the study by Guillaume Piessen et al. [48].

The final model with the comorbidity classes did not include sex, which was traditionally considered

important. However, we found sex differences in different classes, so we speculate that when using LCA method to construct the comorbidity class variable, the potential impact of sex was already included. Therefore, in Lasso screening, comorbidity class was selected and sex was removed.

Our study has certain limitations. First, This study was a retrospective analysis of prospectively collected trial data, which may be subject to selection bias, as well as information bias caused by incomplete or inaccurate data recording. Second, the study may not cover all types of comorbidities, and we were only aware if a patient had a certain type of comorbidity, without knowing the severity of the comorbidity (such as stage 1 hypertension or stage 2 hypertension, well-controlled diabetes or poorly-controlled diabetes). If we can collect more detailed information about comorbidities in the future, it will help improve the accuracy of the model. Third, although there is no external validation, this study still has specific value based on multi-center cooperation. This study included patient from 45 hospitals in 20 provinces, which is representative in exploring the clustering rules of comorbidities and fitting high-performance prognosis prediction models. In the future, to enhance the reliability and reproducibility of our research, we plan to actively pursue external validation and conduct further validation studies using public datasets. Additionally, we aim to develop a user-friendly web application that will allow doctors to quickly determine the classification of patients' comorbidities by simply inputting relevant information, thereby providing more effective support for clinical practice.

Conclusion

We identified two classes of comorbidities in patients with GTC that were associated with overall survival: the first class with no or few comorbidities, and the second class with a high presence and combination of comorbidities. We found that the combination of different comorbidities plays a vital role in the prognosis of these cancers.

Abbreviations

EC	Esophagus cancer
GC	Gastric cancer
CRC	Colorectal cancer
INSCOC	The Investigation on Nutritional Status and its Clinical Outcomes of Common Cancers in China
ICD-10	The International Classification of Diseases, Tenth Revision
NRS2002	Nutritional risk screening 2002
KPS	Karnofsky Performance Scale
BMI	Body Mass Index
NLR	The neutrophil-lymphocyte ratio
PLR	The platelet-lymphocyte ratio
SII	The systemic immunoinflammatory index
PNI	The prognostic nutritional index
LCA	Latent Class Analysis
AIC	Akaike information criterion
LMRT	Lo-Mendell-Rubin adjusted likelihood ratio test
RF	Relative frequency for smallest class
LASSO	Least Absolute Shrinkage and Selection Operator
ROC	Receiver operating characteristic curve

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-025-13517-1.

Supplementary File 1.
Supplementary Table 1.
Supplementary Table 2.
Supplementary Table 3.
Supplementary Table 4.
Supplementary Table 5.
Supplementary Figure 1.
Supplementary Figure 2.
Supplementary Figure 3.
Supplementary Figure 4.

Acknowledgements

The authors thank all the staff of China Cancer Nutrition Database for their hard work.

Authors' contributions

LG: conceptualization, methodology, writing (original draft preparation, review & editing); TY, SG: conceptualization, supervision, writing (review & editing); JC, WL, ZG, HX, MW, SL, QY, WH, LZ, JChen, XW, QZ, HL: data curation; HS, YB, HH: conceptualization, supervision, methodology, writing (review & editing).All authors read and approved the final manuscript.

Funding

This work was supported by the National Key Research and Development Program (2022YFC2009600, 2022YFC2009601), the Shanxi Provincial Basic Research Program (202303021211202), and Doctoral Startup Fund by Shanxi Medical University (XD2104).

Data availability

The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

The data was extracted from the Investigation on Nutritional Status and its Clinical Outcomes of Common Cancers (INSCOC) project in China (chictr.org.cn: ChiCTR1800020329). The study design was approved by the local ethics committees of each participating hospital, and the project was conducted in accordance with the updated Declaration of Helsinki. Before participating in the study, all recruited patients signed a written agreement form authorizing the use of their data.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Gastrointestinal Surgery, First Hospital of Shanxi Medical University, The First Clinical Medical College of Shanxi Medical University, Taiyuan, Shanxi 030001, China. [2]Department of Gastrointestinal Surgery, Center of Clinical Epidemiology and Evidence Based Medicine, First Hospital of Shanxi

Medical University, Center of Clinical Epidemiology and Evidence Based Medicine, Shanxi Medical University, Taiyuan, Shanxi 030001, China. [3]Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Tianjin's Clinical Research Center for Cancer, Tianjin 300060, China. [4]Cancer Center, The First Hospital of Jilin University, Changchun 130021, China. [5]Fujian Cancer Hospital, Fuzhou 350000, China. [6]Department of Clinical Nutrition, Daping Hospital, Army Medical University (Third Military Medical University), Chongqing 400042, China. [7]Department of Clinical Nutrition, The First Affiliated Hospital of Kunming Medical University, Kunming 650500, China. [8]Department of Medical Oncology, Anhui Provincial Cancer Hospital, Hefei 230031, China. [9]The Second Affiliated Hospital of Zhejiang Chinese Medical University, Xinhua Hospital of Zhejiang Province, Hangzhou, Zhejiang 310005, China. [10]Department of Clinical Nutrition, Sichuan University West China Hospital, Chengdu 610041, China. [11]Department of Clinical Nutrition, Yunnan Cancer Hospital, The Third Affiliated Hospital of Kunming Medical University, Kunming, Yunnan 650118, China. [12]Department of Gastrointestinal Surgery, First Affiliated Hospital of Guangxi Medical University, Nanning 530021, China. [13]Xijing Hospital, Air Force Medical University, Xi'an 750102, China. [14]Tianjin Medical University Cancer Institute and Hospital, Tianjin 300060, China. [15]Department of Gastrointestinal Surgery and Department of Clinical Nutrition, Beijing Shijitan Hospital, Capital Medical University, Haidian District, Beijing, Shanxi 100038, China. [16]Peking Union Medical College Hospital, Beijing 100730, China. [17]Department of Nutrition and Food Hygiene, School of Public Health, Shanxi Medical University, Taiyuan, Shanxi 030001, China.

Received: 15 July 2024 Accepted: 14 January 2025
Published online: 08 February 2025

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer statistics 2020: GLOBOCAN estimates of incidence and Mortality Worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71(3):209–49. https://doi.org/10.3322/caac.21660.
2. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2021. CA Cancer J Clin. 2021;71(1):7–33. https://doi.org/10.3322/caac.21654.
3. Jin H, Pinheiro PS, Callahan KE, Altekruse SF. Examining the gastric cancer survival gap between asians and whites in the United States. Gastric Cancer off J Int Gastric Cancer Assoc Jpn Gastric Cancer Assoc. 2017;20(4):573–82. https://doi.org/10.1007/s10120-016-0667-4.
4. Backemar L, Lagergren P, Johar A, Lagergren J. Impact of co-morbidity on mortality after oesophageal cancer surgery. Br J Surg. 2015;102(9):1097–105. https://doi.org/10.1002/bjs.9854.
5. Zhang X, Wang X, Wang M, Gu J, Guo H, Yang Y, et al. Effect of comorbidity assessed by the Charlson Comorbidity Index on the length of stay, costs, and mortality among colorectal cancer patients undergoing colorectal surgery. Curr Med Res Opin. 2023;39(2):187–95. https://doi.org/10.1080/03007995.2022.2139053.
6. Kubo Y, Tanaka K, Yamasaki M, Yamashita K, Makino T, Saito T, et al. Influences of the Charlson comorbidity index and nutrition status on prognosis after esophageal cancer surgery. Ann Surg Oncol. 2021;28(12):7173–82. https://doi.org/10.1245/s10434-021-09779-1.
7. Cao W, Liu S, Wang S, Wang S, Song Y, He Y. Effect of Charlson comorbidity index and treatment strategy on survival of elderly patients after endoscopic submucosal dissection for gastric adenocarcinoma: a multicenter retrospective study. Front Public Health. 2022;9:803113. https://doi.org/10.3389/fpubh.2021.803113.
8. Yoon SJ, Kim EJ, Seo HJ, Oh IH. The association between Charlson comorbidity index and the medical care cost of cancer: a retrospective study. BioMed Res Int. 2015;2015:259341. https://doi.org/10.1155/2015/259341.
9. Lemmens VEPP, Van Halteren AH, Janssen-Heijnen MLG, Vreugdenhil G, Van Repelaer OJ, Coebergh JWW. Adjuvant treatment for elderly patients with stage III colon cancer in the southern Netherlands is affected by socioeconomic status, gender, and comorbidity. Ann Oncol. 2005;16(5):767–72. https://doi.org/10.1093/annonc/mdi159.
10. Sogaard M, Thomsen W, Bossen R, Sørensen HT, Nørgaard H. M. The impact of comorbidity on cancer survival: a review. Clin Epidemiol Published Online November 2013:3. https://doi.org/10.2147/CLEP.S47150.
11. Zeber JE, Copeland LA, Hosek BJ, Karnad AB, Lawrence VA, Sanchez-Reilly SE. Cancer rates, medical comorbidities, and treatment modalities in the oldest patients. Crit Rev Oncol Hematol. 2008;67(3):237–42. https://doi.org/10.1016/j.critrevonc.2008.02.002.
12. Janssen-Heijnen MLG, Houterman S, Lemmens VEPP, Louwman MWJ, Maas HAAM, Coebergh JWW. Prognostic impact of increasing age and co-morbidity in cancer patients: a population-based approach. Crit Rev Oncol Hematol. 2005;55(3):231–40. https://doi.org/10.1016/j.critrevonc.2005.04.008.
13. Keating NL, Landrum MB, Klabunde CN, Fletcher RH, Rogers SO, Doucette WR, et al. Adjuvant chemotherapy for stage III colon cancer: do physicians agree about the importance of patient age and comorbidity? J Clin Oncol off J Am Soc Clin Oncol. 2008;26(15):2532–7. https://doi.org/10.1200/JCO.2007.15.9434.
14. Shack LG, Rachet B, Williams EMI, Northover JMA, Coleman MP. Does the timing of comorbidity affect colorectal cancer survival? A population based study. Postgrad Med J. 2010;86(1012):73–8. https://doi.org/10.1136/pgmj.2009.084566.
15. Xu H, Song C, Yin L, Wang C, Fu Z, Guo Z, et al. Extension protocol for the investigation on nutrition status and clinical outcome of patients with common cancers in China (INSCOC) study: 2021 update. Precis Nutr. 2022;1(2):e00014. https://doi.org/10.1097/PN9.0000000000000014.
16. Investigation on Nutrition Status and its Clinical Outcome of Common Cancers. https://www.chictr.org.cn/showprojEN.html?proj=31813. Accessed 20 Dec 2024.
17. Muthén B, Muthén LK. Integrating person-centered and variable-centered analyses: growth mixture modeling with latent trajectory classes. Alcohol Clin Exp Res. 2000;24(6):882–91.
18. Lanza ST. Latent class analysis for developmental research. Child Dev Perspect. 2016;10(1):59–64. https://doi.org/10.1111/cdep.12163.
19. Kim SY. Determining the number of latent classes in single- and multi-phase growth mixture models. Struct Equ Model Multidiscip J. 2014;21(2):263–79. https://doi.org/10.1080/10705511.2014.882690.
20. Muthen LK, Muthen BO. Mplus user's guide. 8th ed. Los Angeles, CA: Muthen&Muthen; 2020.
21. IBM Corp. IBM SPSS statistics for windows, version 29.0.1. Armonk. NY: IBM Corp; 2021.
22. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2023. URL https://www.R-project.org/.
23. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010;33(1):1–22. https://doi.org/10.18637/R/R/glmnet.
24. Harrell FE. rms: Regression modeling strategies [R package version 5.1-6]. 2021. https://CRAN.R-project.org/package=rms.
25. Wagenmakers T, Van Der Putten K. VIM: Variable importance measures [R package version 1.0–5]. 2014. https://cran.r-project.org/package=VIM.
26. Therneau TM. A package for survival analysis in R. R package version X.Y-Z. 2020. URL https://CRAN.R-project.org/package=survival.
27. Jankowski M, van Calster B. Dcurves: an R package for analyzing decision curves. J Stat Softw. 2017;83(5):1–23. https://doi.org/10.18637/JSSO.V083.I05.
28. Moon KW, Kim SY. autoReg: Automatic linear and logistic regression and survival analysis (version 0.3.3). J Stat Softw. 2023;95(1):1–25. https://doi.org/10.32614/CRAN.package.autoReg.
29. Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer-Verlag; 2009. 978-3-319-24277-4 https://ggplot2.tidyverse.org.
30. Achuthan A, Nadarajah S. timeROC: time-dependent ROC curve analysis (version 1.3.0). R package. 2018. https://CRAN.R-project.org/package=timeROC.
31. Arnold M, Abnet CC, Neale RE, Vignat J, Giovannucci EL, McGlynn KA, et al. Global burden of 5 major types of gastrointestinal Cancer. Gastroenterology. 2020;159(1):335-e34915. https://doi.org/10.1053/j.gastro.2020.02.068.
32. Kim DY, Park YK, Joo JK, Ryu SY, Kim YJ, Kim SK, et al. Clinicopathological characteristics of signet ring cell carcinoma of the stomach. ANZ J Surg. 2004;74(12):1060–4. https://doi.org/10.1111/j.1445-1433.2004.03268.x.
33. Zanders MMJ, Vissers PaJ, Haak HR, van de Poll-Franse LV. Colorectal cancer, diabetes and survival: epidemiological insights. Diabetes Metab. 2014;40(2):120–7. https://doi.org/10.1016/j.diabet.2013.12.007.
34. Giovannucci E, Harlan DM, Archer MC, Bergenstal RM, Gapstur SM, Habel LA, et al. Diabetes and cancer: a consensus report. CA Cancer J Clin. 2010;60(4):207–21. https://doi.org/10.3322/caac.20078.

35. Gronnier C, Messager M, Robb WB, Thiebot T, Louis D, Luc G, et al. Is the negative prognostic impact of signet ring cell histology maintained in early gastric adenocarcinoma? Surgery. 2013;154(5):1093–9. https://doi.org/10.1016/j.surg.2013.05.020.
36. Baré M, Montón C, Mora L, Redondo M, Pont M, Escobar A, et al. COPD is a clear risk factor for increased use of resources and adverse outcomes in patients undergoing intervention for colorectal cancer: a nationwide study in Spain. Int J Chron Obstruct Pulmon Dis. 2017;12:1233–41. https://doi.org/10.2147/COPD.S130377.
37. Choi Y, Kim N, Kim KW, Jo HH, Park J, Yoon H, et al. Sex-based differences in histology, staging, and prognosis among 2983 gastric cancer surgery patients. World J Gastroenterol. 2022;28(9):933–47. https://doi.org/10.3748/wjg.v28.i9.933.
38. Zheng X, Zhang A, Xiao Y, Guo K, Sun L, Ruan S, et al. What causes death in esophageal cancer patients other than the cancer itself: a large population-based analysis. J Cancer. 2022;13(13):3485–94. https://doi.org/10.7150/jca.78004.
39. Zheng Y, Huang Y, Zheng X, Peng J, Chen Y, Yu K, et al. Deaths from COPD in patients with cancer: a population-based study. Aging. 2021;13(9):12641–59. https://doi.org/10.18632/aging.202939.
40. Karlin NJ, Buras MR, Kosiorek HE, Verona PM, Cook CB. Glycemic control and survival of patients with coexisting diabetes mellitus and gastric or esophageal cancer. Future Sci OA. 2019;5(6):FSO397. https://doi.org/10.2144/fsoa-2019-0038.
41. Manohar H, Potter AS, Koutroumpakis E, Deswal A, Palaskas NL. Can we mitigate coronary heart disease risk in patients with cancer? Curr Atheroscler Rep. 2022;24(8):599–606. https://doi.org/10.1007/s11883-022-01035-5.
42. Hopkins BD, Goncalves MD, Cantley LC. Obesity and cancer mechanisms: cancer metabolism. J Clin Oncol. 2016;34(35):4277–83. https://doi.org/10.1200/JCO.2016.67.9712.
43. Ha TK, An JY, Youn HK, Noh JH, Sohn TS, Kim S. Indication for endoscopic mucosal resection in early signet ring cell gastric cancer. Ann Surg Oncol. 2008;15(2):508–13. https://doi.org/10.1245/s10434-007-9660-9.
44. Kunisaki C, Shimada H, Nomura M, Matsuda G, Otsuka Y, Akiyama H. Therapeutic strategy for signet ring cell carcinoma of the stomach. Br J Surg. 2004;91(10):1319–24. https://doi.org/10.1002/bjs.4637.
45. Li C, Kim S, Lai JF, Hyung WJ, Choi WH, Choi SH, et al. Advanced gastric carcinoma with signet ring cell histology. Oncology. 2007;72(1–2):64–8. https://doi.org/10.1159/000111096.
46. Kwon KJ, Shim KN, Song EM, Choi JY, Kim SE, Jung HK, et al. Clinicopathological characteristics and prognosis of signet ring cell carcinoma of the stomach. Gastric Cancer off J Int Gastric Cancer Assoc Jpn Gastric Cancer Assoc. 2014;17(1):43–53. https://doi.org/10.1007/s10120-013-0234-1.
47. Kim JP, Kim SC, Yang HK. Prognostic significance of signet ring cell carcinoma of the stomach. Surg Oncol. 1994;3(4):221–7. https://doi.org/10.1016/0960-7404(94)90037-x.
48. Piessen G, Messager M, Leteurtre E, Jean-Pierre T, Mariette C. Signet ring cell histology is an independent predictor of poor prognosis in gastric adenocarcinoma regardless of tumoral clinical presentation. Ann Surg. 2009;250(6):878–87. https://doi.org/10.1097/SLA.0b013e3181b21c7b.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.