



## OPEN ACCESS

## EDITED BY

Alireza Sadjadi,  
Tehran University of Medical Sciences, Iran

## REVIEWED BY

Aimin Jiang,  
Shandong Tumor Hospital, China  
Xuhua Hu,  
Fourth Hospital of Hebei Medical University,  
China

## \*CORRESPONDENCE

Hai Li  
✉ 13988333828@163.com  
Changhua Zhang  
✉ zhangchanghua@sysush.com  
Yulong He  
✉ heyulong@sysush.com

†These authors have contributed equally to  
this work

RECEIVED 07 December 2023

ACCEPTED 21 October 2024

PUBLISHED 11 November 2024

## CITATION

Huang Z, Chen S, Yin S, Shi Z, Gu L, Li L,  
Yin H, Huang Z, Li B, Chen X, Yang Y, Wang Z,  
Li H, Zhang C and He Y (2024) Development  
and validation of a nomogram for predicting  
the risk of developing gastric cancer based on  
a questionnaire: a cross-sectional study.  
*Front. Oncol.* 14:1351967.  
doi: 10.3389/fonc.2024.1351967

## COPYRIGHT

© 2024 Huang, Chen, Yin, Shi, Gu, Li, Yin,  
Huang, Li, Chen, Yang, Wang, Li, Zhang and  
He. This is an open-access article distributed  
under the terms of the [Creative Commons  
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,  
distribution or reproduction in other forums  
is permitted, provided the original author(s)  
and the copyright owner(s) are credited and  
that the original publication in this journal is  
cited, in accordance with accepted academic  
practice. No use, distribution or reproduction  
is permitted which does not comply with  
these terms.

# Development and validation of a nomogram for predicting the risk of developing gastric cancer based on a questionnaire: a cross-sectional study

Zhangsen Huang<sup>1†</sup>, Songyao Chen<sup>1†</sup>, Songcheng Yin<sup>1,2†</sup>,  
Zhaowen Shi<sup>3†</sup>, Liang Gu<sup>1</sup>, Liang Li<sup>1</sup>, Haofan Yin<sup>1</sup>,  
Zhijian Huang<sup>1</sup>, Bo Li<sup>1,2</sup>, Xin Chen<sup>3</sup>, Yilin Yang<sup>3</sup>, Zhengli Wang<sup>3</sup>,  
Hai Li<sup>3\*</sup>, Changhua Zhang<sup>1,2\*</sup> and Yulong He<sup>1,2\*</sup>

<sup>1</sup>Digestive Medicine Center, The Seventh Affiliated Hospital of Sun Yat-sen University, Shenzhen, China, <sup>2</sup>Guangdong Provincial Key Laboratory of Digestive Cancer Research, Digestive Diseases Center, The Seventh Affiliated Hospital of Sun Yat-Sen University, Shenzhen, Guangdong, China, <sup>3</sup>General Surgery, Fengqing People's Hospital, Lincang, China

**Background:** Detection of gastric cancer (GC) at early stages is an effective strategy for decreasing mortality. This study aimed to construct a prediction nomogram based on a questionnaire to assess the risk of developing GC.

**Methods:** Our study comprised a total of 4379 participants (2326 participants from outpatient at Fengqing People's Hospital were considered for model development and internal validation, and 2053 participants from outpatients at the endoscopy center at the Seventh Affiliated Hospital of Sun Yat-Sen University were considered for independent external validation) and gastric mucosa status was determined by endoscopy and biopsies. The eligible participants in development cohort from Fengqing people's Hospital were randomly separated into a training cohort (n=1629, 70.0%) and an internal validation cohort (n=697, 30.0%). The relevant features were selected by a least absolute shrinkage and selection operator (LASSO), and the ensuing features were evaluated through multivariable logistic regression analysis. Subsequently, the variables were selected to construct a prediction nomogram. The discriminative ability and predictive accuracy of the nomogram were evaluated by the C-index and calibration plot, respectively. Decision curve analysis (DCA) curves were used for the assessment of clinical benefit of the model. This model was developed to estimate the risk of developing neoplastic lesions according to the "transparent reporting of a multivariable prediction model for individual prognosis or diagnosis" (TRIPOD) statement.

**Results:** Six variables, including age, sex, alcohol consumption, cigarette smoking, education level, and Hp infection status, were independent risk factors for the development of neoplastic lesions. Thus, these variables were incorporated into the final nomogram. The AUC of the nomogram were 0.701, 0.657 and 0.699 in the training, internal validation, and external validation cohorts, respectively. The calibration curve showed that the nomogram was in

good agreement with the observed outcomes. Compared to treatment of all patients or none, our nomogram showed a notably higher clinical benefit.

**Conclusion:** This nomogram proved to be a convenient, cost-effective tool to effectively predict an individual's risk of developing neoplastic lesions, and it can act as a prescreening tool before gastroscopy.

#### KEYWORDS

gastric cancer, nomogram, risk factor, cancer screening, questionnaire, prediction model

## Introduction

Gastric cancer (GC) is a common malignancy worldwide, ranking fifth in terms of both incidence and cancer-related deaths among malignancies. In 2022, there were over 968,000 new cases and 659,000 deaths globally (1). In China, GC ranks fifth in incidence and third in mortality among malignancies, accounting for 37% of the global incidence and 39.4% of the global mortality (2). The high incidence and mortality rates of GC have posed an enormous burden on China's healthcare system. The high mortality rate is largely due to low early detection rate, and thus, early detection and treatment play a pivotal role in reducing GC-related deaths. When the diagnosis is made at an early stage, the surgical treatment results in the 5-year survival rate can exceed 90%, even achieving complete remission (3). Indeed, the National Upper Gastrointestinal Cancer Early Detection program, composing GC and esophageal cancer screening program, has achieved remarkable progress in the early detection of GC in China (4), the 5-year survival rate for GC ranges from 35% to 44% in China (3, 5). Fengqing County, located in the southwest region of Yunnan Province, has a high incidence of GC and related mortality. Based on data obtained from Fengqing's Centers for Disease Control and Prevention, GC is the first most common cancer in Fengqing County, and the 5-year survival rate is only 25.7%, which is lower than the national average survival rate (Supplementary Figure S1). Thus, it is imperative to prioritize prevention and timely diagnosis to improve the prognosis of GC.

Currently, upper gastrointestinal endoscopic examination serves as the gold standard for screening and diagnosing GC. However, gastroscopy-based screening is inefficient and impracticable due to the large population that needs screening in China. Hence, a risk stratification method is required as a prescreening tool before gastroscopy. Recently, a risk scoring system was proposed to identify high-risk individuals for GC, who will be subject to endoscopic screening (6). This system incorporates seven predictors, including pickled food

consumption, fried food consumption, age, sex, gastrin-17 (G-17), pepsinogen I/II ratio (PGR), and *Helicobacter pylori* (*H. pylori*, Hp) infection (6). This risk scoring system can identify 69.6% of GC in an external validation cohort. However, including serum indicators predictors would substantially increase the cost of risk evaluation and restrict its generalizability, especially in economically disadvantaged regions such as Yunnan Province. Hence, there is an urgent need to establish a non-invasive, simple-to-use, and accurate risk prediction model that can act as a prescreening tool for identifying individuals with a high risk of GC. Such a model would aid in deciding whether endoscopic examination is necessary.

GC is typically a multifactorial and multistep pathological progression. It starts with a series of gastric mucosal changes in the normal mucosa, which progress into non-atrophic gastritis (no-CAG), chronic atrophic gastritis (CAG), intestinal metaplasia (IM), gastric intraepithelial neoplasia (GIN) and ultimately carcinoma (as proposed by Correa's hypothesis) (7, 8). According to Japanese data, the 5-year cumulative GC incidence ranges from 1.9% to 10% in CAG and from 5.3% to 9.8% in IM (9). Furthermore, the risk of GC occurrence in cases of biopsy-detected GIN are reportedly to be 2.8% to 11.5% for low-grade intraepithelial neoplasia (LGIN) and 10% to 68.8% for high-grade intraepithelial neoplasia (HGIN) (10–13). Thus, enhancing the diagnosis rate of GIN can improve the prognosis. The stage of the GIN is crucial in the progression of GC, as it offers the best chance to avoid GC development. Identification and administration of appropriate intervention or treatment to patients with GIN can reduce the incidence of GC and improve patients' outcomes.

Nomogram, which is a visual graphic tool that calculates each individual's probability of clinical events considering the preweight value of each factor, has been widely used to predict the risk and prognosis of various cancers in recent years (14, 15). In this study, we aimed to establish and evaluate a convenient and practical prediction model to assess the risk of developing neoplastic lesions, including GIN and carcinoma. And this model aims to identify a subgroup of individuals at high risk of GC among Chinese population for further diagnostic gastroscopy.

## Materials and methods

### Ethics approval

The study was approved by the Ethics Committee of Fengqing People’s Hospital and the Seventh Affiliated Hospital of Sun Yat-Sen University (KY-2021-105-01), and had been performed in line with the ethical standards in the Declaration of Helsinki. Written informed consents were obtained from all participants in this study.

### Study population

A total of 4379 participants underwent upper gastrointestinal endoscopic examination at the endoscopy center at Fengqing people’s Hospital and the Seventh Affiliated Hospital of Sun Yat-Sen University were included in this study between September 2019 and March 2024, 2326 participants were from outpatients at the endoscopy center at Fengqing people’s Hospital and 2053 participants were from outpatients at the endoscopy center at the Seventh Affiliated Hospital of Sun Yat-Sen University. Participants in the development cohort from Fengqing people’s Hospital were randomly divided into a developing set [n=1629, 70.0%] and an internal validation set [n=697, 30.0%], the participants from outpatients at the endoscopy center at the Seventh Affiliated Hospital of Sun Yat-Sen University were used for external validation [n=2053] (Figure 1). All participants had completed of

upper gastrointestinal endoscopy with valid examination results and were reported as normal gastric mucosa, no-CAG, CAG, IM, LGIN, and HGIN, or carcinoma. These participants were further divided into two groups: non-neoplastic lesions (normal gastric mucosa, no-CAG, CAG, and IM) and neoplastic lesions (LGIN, HGIN, and carcinoma) according to whether the lesion is negative for neoplasia.

### Inclusion and exclusion criteria

The eligibility requirements were as follows: 1) Aged between 35 and 85 years; 2) No history of cancer, mental disorders, or any contraindication to endoscopy; and 3) Had completed of upper gastrointestinal endoscopy with valid examination results. Participants who with a history of cancer were excluded from this study. In addition, we excluded participants who were diagnosed with other malignancies or incomplete information.

The lower age limit of 35 years was used in present study in accordance with the criteria established by the Wuwei Cohort (16), and identifying individuals with GIN at an earlier age could facilitate the implementation of effective preventive strategies. There is no upper age limit in GC screening programs in Japan (17, 18). The upper age limit of 85 years was used in present study as evidenced by data indicating that the incidence rate among individuals aged 75–84 years accounts for more than one-fifth (75,724/358,672) in China (2).

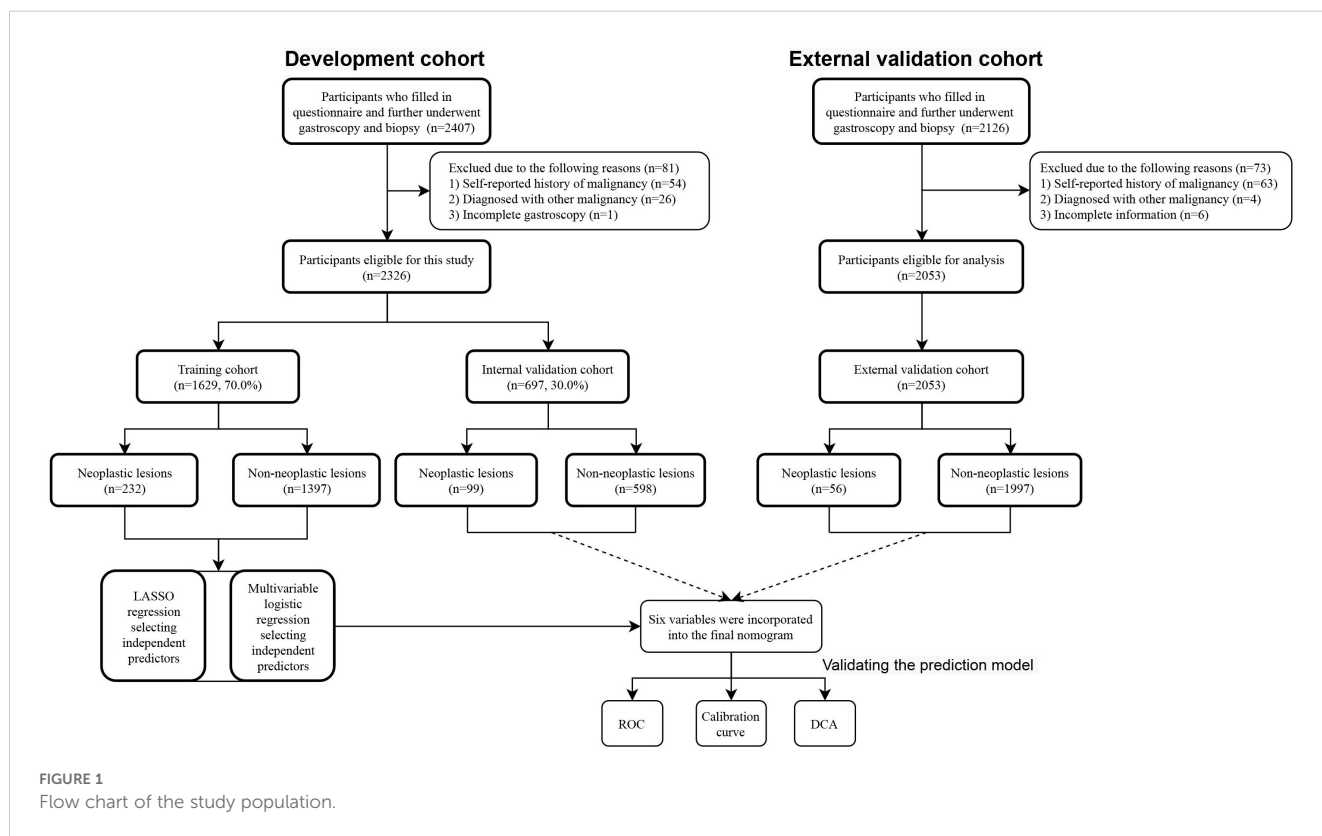


FIGURE 1  
Flow chart of the study population.

## Data collection

A prevalidated self-report questionnaire was used in this study, and the interviews were conducted face-to-face by highly trained investigators. All variables were questionnaire-based predictors that met any one of the following criteria: 1) biologically plausible in predicting GC risk or 2) reported in at least two published risk prediction studies. The questionnaire included baseline demographic information (age, sex, ethnic group, sleep quality [sleep quality defined as satisfaction with one's sleep experience, including bedtime, times of wake up in the night, frequency difficulty falling asleep, frequency difficulty getting back to sleep after waking up, and frequency wake up feeling tired and worn out, according to a published paper (19)], education, occupation, marital status, and household income, body weight and height), lifestyle (cigarette smoking [smoking more than one cigarette every day for more than 1 year or ever smoked], alcohol consumption [drinking any type of alcohol more than once every week lasting for more than 1 year or ever drank], dietary habits (hot food [hot food means that the temperature of food is over 70 °C], pickled food, fried food, smoked food, spicy food, overnight leftovers, mealtime [mealtime regular was defined as regular time for meals across days and regular meal frequency], red meat, green vegetables and fresh fruit [frequency was defined as more than 3 times/week, frequent fruits and vegetables frequency was defined as more than 3 times/week or >2500 g/week for vegetables and >1250 g/week for fruits], medical history (painkiller, antiplatelet agents, chronic gastritis, gastric ulcer, gastric polyp, hypertrophic gastritis, gastric resection, hyperlipemia, hypertension, diabetes, and pernicious anemia), and family history of GC (FHGC) in first-degree relatives. All endoscopic examinations were performed by well-trained endoscopy doctors at Fengqing People's Hospital and the Seventh Affiliated Hospital of Sun Yat-Sen University. Pathologists then independently read the biopsy slides and made pathological diagnosis of the biopsy specimens. Family income in ten thousand Yuan and it then was categorized into two groups: low (<5), and high (≥5).

## Nomogram construction

To minimize the potential variables, a least absolute shrinkage and selection operator (LASSO) regression was used in this study. The optimal penalization coefficient ( $\lambda$ ) in the LASSO model was identified by 10-fold cross-validation on the training cohort via minimum criterion. With larger penalties, the estimates of weaker factors shrink towards zero, leaving only the strongest predictors in the model. In the present study, the predictive covariates were selected by the minimum ( $\lambda$  min). These variables identified by LASSO regression analysis were selected for further analysis. Multivariable analyses were performed to determine the independent predictors ( $p < 0.05$ ), and the nomogram was developed with the selected independent predictors. Each independent risk factor in the nomogram was assigned a score, and the total score was calculated from the participant data to

predict the probability of developing GC. The LASSO regression was performed by using "glmnet" package in RStudio software, and the risk prediction nomogram for developing GC was performed using the "rms" package in RStudio software. The model development, internal-external validation, and reporting were performed following the "transparent reporting of a multivariable prediction model for individual prognosis or diagnosis" (TRIPOD) guidelines (20).

## Statistical analysis

Continuous variables were presented as the mean  $\pm$  standard deviation (SD). Categorical variables were expressed as frequencies. Differences between the two groups for continuous variables were analyzed by the student's test or Mann-Whitney U test, and the Pearson's Chi-Square ( $\chi^2$ ) test was used for categorical variables. The independent predictors for the development of GC were assessed by multivariable logistic regression analyses. A nomogram was then formulated using the rms package in RStudio software, based on the results of the multivariable logistic regression analyses. The discrimination power of the nomogram was assessed by Harrell's concordance index (C-index). A calibration plot was used to describe the consistency between the predicted and the actual probability of the occurrence of neoplastic lesions. A decision curve analysis (DCA) graph was generated to evaluate the clinical effectiveness of the model. All statistical analyses were performed using RStudio software (version 4.2.1). Statistically significant in all statistical analyses was defined as  $p < 0.05$  in a two-tailed test.

## Results

### Participant's characteristics

A total of 4379 participants underwent upper gastrointestinal endoscopic examination at the endoscopy center were included in this study between September 2019 and March 2024, 2326 participants were from outpatients at the endoscopy center at Fengqing people's Hospital and 2053 participants were from outpatients at the endoscopy center at the Seventh Affiliated Hospital of Sun Yat-Sen University. A total of 2326 participants in the development cohort, including 1690 cases of no-CAG, 132 cases of CAG, 173 cases of IM, 211 cases of GIN and 120 cases of GC, were recruited in the study, and the participants were divided into two groups: non-neoplastic lesions ( $n=1995$ ; 1690 participants with normal gastric mucosa or no-CAG, 132 participants with CAG, and 173 participants with IM) and neoplastic lesions ( $n=331$ ; 211 participants with GIN and 120 participants with GC). A total of 2053 participants, including 1488 cases of no-CAG, 80 cases of CAG, 429 cases of IM, 17 cases of GIN and 39 cases of GC, were recruited from November 2021 to March 2024 as the external validation cohort. A flow chart showing the process of including participants in the study is depicted in Figure 1. In the development cohort, there were 1160 males (49.9%) and 1166 females (50.1%),

TABLE 1 Characteristics of the participants who enrolled in our study.

Variables*	Training Cohort (n = 1629)			Internal Validation Cohort (n = 697)			External Validation Cohort (n = 2053)		
	Neoplastic lesions (n = 232)	Non-neoplastic lesions (n = 1397)	p-value	Neoplastic lesions (n = 99)	Non-neoplastic lesions (n = 598)	p-value	Neoplastic lesions (n = 56)	Non-neoplastic lesions (n = 1997)	p-value
<b>Age, years, n (%)</b>			<b>0.000</b>			<b>0.039</b>			<b>0.000</b>
<50	45 (19.40%)	440 (31.50%)		21 (21.21%)	159 (26.59%)		13 (23.21%)	822 (41.16%)	
50-59	81 (34.91%)	560 (40.09%)		33 (33.33%)	255 (42.64%)		15 (26.79%)	709 (35.50%)	
60-69	77 (33.19%)	302 (21.62%)		35 (35.35%)	146 (24.41%)		12 (21.43%)	342 (17.13%)	
>69	29 (12.50%)	95 (6.80%)		10 (10.10%)	38 (6.35%)		16 (28.57%)	124 (6.21%)	
<b>Sex, n (%)</b>			<b>0.000</b>			<b>0.000</b>			<b>0.001</b>
Female	67 (28.88%)	749 (53.61%)		29 (29.29%)	321 (53.68%)		14 (25.00%)	949 (47.52%)	
Male	165 (71.12%)	648 (46.39%)		70 (70.71%)	277 (46.32%)		42 (75.00%)	1048 (52.48%)	
<b>Ethnic group, n (%)</b>			<b>0.095</b>			<b>0.160</b>			<b>0.047</b>
Han	165 (71.12%)	1068 (76.45%)		80 (80.81%)	440 (73.58%)		54 (96.43%)	1986 (99.45%)	
Others	67 (28.88%)	329 (23.55%)		19 (19.19%)	158 (26.42%)		2 (3.57%)	11 (0.55%)	
<b>Painkiller intake, n (%)</b>			<b>0.627</b>			<b>0.603</b>			<b>0.118</b>
No	171 (73.71%)	1054 (75.45%)		72 (72.73%)	453 (75.75%)		54 (96.43%)	1782 (89.23%)	
Yes	61 (26.29%)	343 (24.55%)		27 (27.27%)	145 (24.25%)		2 (3.57%)	215 (10.77%)	
<b>Antiplatelet agents intake, n (%)</b>			<b>0.320</b>			<b>0.913</b>			<b>0.784</b>
No	203 (87.50%)	1256 (89.91%)		90 (90.91%)	538 (89.97%)		52 (92.86%)	1866 (93.44%)	
Yes	29 (12.50%)	141 (10.09%)		9 (9.09%)	60 (10.03%)		4 (7.14%)	131 (6.56%)	
<b>Sleep quality, n (%)</b>			<b>0.969</b>			<b>0.950</b>			<b>0.667</b>
Adequate	123 (53.02%)	753 (53.90%)		55 (55.56%)	324 (54.18%)		32 (57.14%)	1010 (50.58%)	
Good	78 (33.62%)	461 (33.00%)		31 (31.31%)	189 (31.61%)		16 (28.57%)	662 (33.15%)	
Poor	31 (13.36%)	183 (13.10%)		13 (13.13%)	85 (14.21%)		8 (14.29%)	325 (16.27%)	
<b>Pickled food intake, n (%)</b>			<b>0.562</b>			<b>0.516</b>			<b>0.348</b>
Not frequent	178 (76.72%)	1099 (78.67%)		78 (78.79%)	491 (82.11%)		52 (92.86%)	1901 (95.19%)	
Frequent	54 (23.28%)	298 (21.33%)		21 (21.21%)	107 (17.89%)		4 (7.14%)	96 (4.81%)	
<b>Smoked food intake, n (%)</b>			<b>0.009</b>			<b>0.205</b>			<b>1.000</b>
Not frequent	180 (77.59%)	1183 (84.68%)		81 (81.82%)	521 (87.12%)		56 (100.00%)	1971 (98.70%)	
Frequent	52 (22.41%)	214 (15.32%)		18 (18.18%)	77 (12.88%)		0 (0.00%)	26 (1.30%)	
<b>History of stomach disease, n (%)**</b>			<b>0.449</b>			<b>0.673</b>			<b>0.266</b>
No	178 (76.72%)	1106 (79.17%)		82 (82.83%)	481 (80.43%)		40 (71.43%)	1552 (77.72%)	
Yes	54 (23.28%)	291 (20.83%)		17 (17.17%)	117 (19.57%)		16 (28.57%)	445 (22.28%)	
<b>FHGC, n (%)***</b>			<b>0.450</b>			<b>0.100</b>			<b>1.000</b>
No	220 (94.83%)	1343 (96.13%)		92 (92.93%)	577 (96.49%)		53 (94.64%)	1883 (94.29%)	
Yes	12 (5.17%)	54 (3.87%)		7 (7.07%)	21 (3.51%)		3 (5.36%)	114 (5.71%)	
<b>Alcohol consumption, n (%)</b>			<b>0.000</b>			<b>0.261</b>			<b>0.512</b>
No	146 (62.93%)	1123 (80.39%)		76 (76.77%)	491 (82.11%)		34 (60.71%)	1106 (55.38%)	

(Continued)

TABLE 1 Continued

Variables*	Training Cohort (n = 1629)			Internal Validation Cohort (n = 697)			External Validation Cohort (n = 2053)		
	Neoplastic lesions (n = 232)	Non-neoplastic lesions (n = 1397)	p-value	Neoplastic lesions (n = 99)	Non-neoplastic lesions (n = 598)	p-value	Neoplastic lesions (n = 56)	Non-neoplastic lesions (n = 1997)	p-value
<b>Alcohol consumption, n (%)</b>			<b>0.000</b>			<b>0.261</b>			<b>0.512</b>
Yes	86 (37.07%)	274 (19.61%)		23 (23.23%)	107 (17.89%)		22 (39.29%)	891 (44.62%)	
<b>Cigarette smoking, n (%)</b>			<b>0.000</b>			<b>0.017</b>			<b>0.016</b>
No	125 (53.88%)	1071 (76.66%)		63 (63.64%)	452 (75.59%)		34 (60.71%)	1513 (75.76%)	
Yes	107 (46.12%)	326 (23.34%)		36 (36.36%)	146 (24.41%)		22 (39.29%)	484 (24.24%)	
<b>Hot food intake, n (%)</b>			<b>0.631</b>			<b>1.000</b>			<b>0.500</b>
Not frequent	168 (72.41%)	1036 (74.16%)		74 (74.75%)	444 (74.25%)		46 (82.14%)	1705 (85.38%)	
Frequent	64 (27.59%)	361 (25.84%)		25 (25.25%)	154 (25.75%)		10 (17.86%)	292 (14.62%)	
<b>Leftover food intake, n (%)</b>			<b>0.019</b>			<b>0.136</b>			<b>0.003</b>
Not frequent	167 (71.98%)	1105 (79.10%)		70 (70.71%)	467 (78.09%)		45 (80.36%)	1850 (92.64%)	
Frequent	65 (28.02%)	292 (20.90%)		29 (29.29%)	131 (21.91%)		11 (19.64%)	147 (7.36%)	
<b>Mealtime, n (%)</b>			<b>0.796</b>			<b>0.471</b>			<b>0.140</b>
Unregular	18 (7.76%)	119 (8.52%)		9 (9.09%)	39 (6.52%)		6 (10.71%)	115 (5.76%)	
Regular	214 (92.24%)	1278 (91.48%)		90 (90.91%)	559 (93.48%)		50 (89.29%)	1882 (94.24%)	
<b>Overeat and overdrink, n (%)</b>			<b>0.456</b>			<b>1.000</b>			<b>0.339</b>
Not frequent	226 (97.41%)	1370 (98.07%)		97 (97.98%)	584 (97.66%)		54 (96.43%)	1955 (97.90%)	
Frequent	6 (2.59%)	27 (1.93%)		2 (2.02%)	14 (2.34%)		2 (3.57%)	42 (2.10%)	
<b>Fresh fruit intake, n (%)</b>			<b>0.713</b>			<b>0.089</b>			<b>0.964</b>
Not frequent	150 (64.66%)	924 (66.14%)		75 (75.76%)	398 (66.56%)		27 (48.21%)	969 (48.52%)	
Frequent	82 (35.34%)	473 (33.86%)		24 (24.24%)	200 (33.44%)		29 (51.79%)	1028 (51.48%)	
<b>Fresh vegetables intake, n (%)</b>			<b>0.047</b>			<b>0.421</b>			<b>0.625</b>
Not frequent	65 (28.02%)	488 (34.93%)		40 (40.40%)	213 (35.62%)		7 (12.50%)	209 (10.47%)	
Frequent	167 (71.98%)	909 (65.07%)		59 (59.60%)	385 (64.38%)		49 (87.50%)	1788 (89.53%)	
<b>Red meat intake, n (%)<sup>#</sup></b>			<b>0.113</b>			<b>0.607</b>			<b>0.463</b>
Not frequent	130 (56.03%)	863 (61.78%)		60 (60.61%)	382 (63.88%)		15 (26.79%)	627 (31.40%)	
Frequent	102 (43.97%)	534 (38.22%)		39 (39.39%)	216 (36.12%)		41 (73.21%)	1370 (68.60%)	
<b>Educational level, n (%)</b>			<b>0.026</b>			<b>0.921</b>			<b>0.032</b>
illiteracy	35 (15.09%)	139 (9.95%)		10 (10.10%)	55 (9.20%)		5 (8.93%)	61 (3.05%)	
literacy	197 (84.91%)	1258 (90.05%)		89 (89.90%)	543 (90.80%)		51 (91.07%)	1936 (96.95%)	
<b>Digestive tract symptom, n (%)<sup>##</sup></b>			<b>0.503</b>			<b>0.913</b>			<b>0.168</b>
No	107 (46.12%)	681 (48.75%)		48 (48.48%)	297 (49.67%)		13 (23.21%)	637 (31.90%)	
Yes	125 (53.88%)	716 (51.25%)		51 (51.52%)	301 (50.33%)		43 (76.79%)	1360 (68.10%)	

(Continued)

TABLE 1 Continued

Variables*	Training Cohort (n = 1629)			Internal Validation Cohort (n = 697)			External Validation Cohort (n = 2053)		
	Neoplastic lesions (n = 232)	Non-neoplastic lesions (n = 1397)	p-value	Neoplastic lesions (n = 99)	Non-neoplastic lesions (n = 598)	p-value	Neoplastic lesions (n = 56)	Non-neoplastic lesions (n = 1997)	p-value
<b>Type of residential area, n (%)</b>			<b>0.764</b>			<b>0.518</b>			<b>0.005</b>
Rural	210 (90.52%)	1252 (89.62%)		92 (92.93%)	540 (90.30%)		16 (28.57%)	299 (14.97%)	
Urban	22 (9.48%)	145 (10.38%)		7 (7.07%)	58 (9.70%)		40 (71.43%)	1698 (85.03%)	
<b>Occupation, n (%)</b>			<b>0.599</b>			<b>0.071</b>			<b>0.012</b>
farmer	184 (79.31%)	1132 (81.03%)		88 (88.89%)	483 (80.77%)		10 (17.86%)	156 (7.81%)	
unfarmer	48 (20.69%)	265 (18.97%)		11 (11.11%)	115 (19.23%)		46 (82.14%)	1841 (92.19%)	
<b>Income, n (%)</b>			<b>1.000</b>			<b>0.559</b>			<b>0.705</b>
High	48 (20.69%)	286 (20.47%)		19 (19.19%)	134 (22.41%)		31 (55.36%)	1156 (57.89%)	
Low	184 (79.31%)	1111 (79.53%)		80 (80.81%)	464 (77.59%)		25 (44.64%)	841 (42.11%)	
<b>Marital status, n (%)###</b>			<b>1.000</b>			<b>0.138</b>			<b>1.000</b>
Married	212 (91.38%)	1278 (91.48%)		87 (87.88%)	555 (92.81%)		51 (91.07%)	1813 (90.79%)	
Unmarried	20 (8.62%)	119 (8.52%)		12 (12.12%)	43 (7.19%)		5 (8.93%)	184 (9.21%)	
<b>Current and Past Manual labour, n (%)</b>			<b>0.462</b>			<b>1.000</b>			<b>0.115</b>
No	30 (12.93%)	210 (15.03%)		12 (12.12%)	71 (11.87%)		35 (62.50%)	1440 (72.11%)	
Yes	202 (87.07%)	1187 (84.97%)		87 (87.88%)	527 (88.13%)		21 (37.50%)	557 (27.89%)	
<b>BMI (kg/m<sup>2</sup>), n (%)<sup>§</sup></b>			<b>0.071</b>			<b>1.000</b>			<b>0.005</b>
<18.5	32 (13.79%)	1097 (78.53%)		9 (9.09%)	52 (8.70%)		7 (12.50%)	72 (3.61%)	
>=18.5	200 (86.21%)	300 (21.47%)		90 (90.91%)	546 (91.30%)		49 (87.50%)	1925 (96.39%)	
<b>Hp, n (%)</b>			<b>0.159</b>			<b>0.222</b>			<b>1.000</b>
Negative	212 (91.38%)	1314 (94.06%)		89 (89.90%)	561 (93.81%)		42 (75.00%)	1510 (75.61%)	
Positive	20 (8.62%)	83 (5.94%)		10 (10.10%)	37 (6.19%)		14 (25.00%)	487 (24.39%)	
<b>Water intake (ml/day), n (%)</b>			<b>0.2817</b>			<b>0.9399</b>			<b>0.590</b>
>1000	39 (16.81%)	194 (13.89%)		16 (16.16%)	102 (17.06%)		15 (26.79%)	601 (30.10%)	
≤1000	193 (83.19%)	1203 (86.11%)		83 (83.84%)	496 (82.94%)		41 (73.21%)	1936 (96.95%)	
<b>Chronic disease, n (%)</b>			<b>0.2385</b>			<b>0.6919</b>			<b>0.504</b>
No	156 (67.24%)	996 (71.30%)		74 (74.75%)	432 (72.24%)		33 (58.93%)	1264 (63.29%)	
Yes	76 (32.76%)	401 (28.70%)		25 (25.25%)	166 (27.76%)		23 (41.07%)	733 (36.71%)	

Data are presented as n for categorical variables, mean (SD) for continuous variables. For variables about eating habits, two categories for frequency of consumption were provided, that is, not frequent (<3 times/week) and frequent (at least 3 times/week).

\*Variables were questionnaire-based predictors which met any one of the criteria below: 1) biologic plausibility in predicting the risk of GC or 2) reported in at least two published risk prediction studies.

\*\*History of stomach diseases includes gastritis, gastric ulcers, hypertrophic gastritis, remnant stomach and gastric polyps.

\*\*\*Family history of gastric cancer (FHGC) is defined as GC cases among first-degree relatives.

#Red meat includes beef, pork and lamb.

##Digestive tract symptom includes heartburn, problems swallowing or passing food along your esophagus and stomach, stomach pain, nausea, and vomiting.

###Marital status classifies as married and unmarried (including single, separated, divorced, widowed, and unmarried patients).

§Body mass index (BMI): weight (kg)/height (m)<sup>2</sup>.

\*p values refer to comparison between low-risk participants and high-risk participants in the univariate analysis.

p values refer to comparison between neoplastic lesions and non-neoplastic lesions in the univariate analysis.

1753 participants were the Han nationality (75.4%) and 573 participants were minority nationality (24.6%), and the mean age was 55.57 ± 9.06 years (range 35–84 years). In the external validation cohort, there were 1090 males (53.1%) and 963 females

(46.9%), 2040 participants were the Han nationality (99.4%) and 13 participants were minority nationality (0.6%), and the mean age was 53.21 ± 9.40 years (range 38–85 years). All data of participants, including baseline information, lifestyle, medical history and

TABLE 2 Characteristics of the training cohort and the internal validation cohort.

Variables	Total participants (n = 2326)	Training Cohort (n = 1629)	Validation Cohort (n = 697)	p-value
Age, years, n (%)				0.174
<50	665 (28.59%)	485 (29.77%)	180 (25.82%)	
50-59	929 (39.94%)	641 (39.35%)	288 (41.32%)	
60-69	560 (24.07%)	379 (23.27%)	181 (25.97%)	
>69	172 (7.40%)	124 (7.61%)	48 (6.89%)	
Sex, n (%)				0.993
Female	1166 (50.13%)	816 (50.09%)	350 (50.22%)	
Male	1160 (49.87%)	813 (49.91%)	347 (49.78%)	
Ethnic group, n (%)				0.614
Han	1753 (75.37%)	1233 (75.69%)	520 (74.61%)	
Others	573 (24.63%)	396 (24.31%)	177 (25.39%)	
Painkiller intake, n (%)				0.992
No	1750 (75.24%)	1225 (75.20%)	525 (75.32%)	
Yes	576 (24.76%)	404 (24.80%)	172 (24.68%)	
Antiplatelet agents intake, n (%)				0.752
No	2087 (89.72%)	1459 (89.56%)	628 (90.10%)	
Yes	239 (10.28%)	170 (10.44%)	69 (9.90%)	
Sleep quality, n (%)				0.708
Adequate	1255 (53.96%)	876 (53.77%)	379 (54.38%)	
Good	759 (32.63%)	539 (33.09%)	220 (31.56%)	
Poor	312 (13.41%)	214 (13.14%)	98 (14.06%)	
Pickled food intake, n (%)				0.086
Not frequent	1846 (79.36%)	1277 (78.39%)	569 (81.64%)	
Frequent	480 (20.64%)	352 (21.61%)	128 (18.36%)	
Smoked food intake, n (%)				0.113
Not frequent	1965 (84.48%)	1363 (83.67%)	602 (86.37%)	
Frequent	361 (15.52%)	266 (16.33%)	95 (13.63%)	
History of stomach disease, n (%)				0.312
No	1847 (79.41%)	1284 (78.82%)	563 (80.77%)	
Yes	479 (20.59%)	345 (21.18%)	134 (19.23%)	
FHGC, n (%)				1.000
No	2232 (95.96%)	1563 (95.95%)	669 (95.98%)	
Yes	94 (4.04%)	66 (4.05%)	28 (4.02%)	
Alcohol consumption, n (%)				0.070
No	1836 (78.93%)	1269 (77.90%)	567 (81.35%)	
Yes	490 (21.07%)	360 (22.10%)	130 (18.65%)	
Cigarette smoking, n (%)				0.854
No	1711 (73.56%)	1196 (73.42%)	515 (73.89%)	

(Continued)



TABLE 2 Continued

Variables	Total participants (n = 2326)	Training Cohort (n = 1629)	Validation Cohort (n = 697)	p-value
Yes	615 (26.44%)	433 (26.58%)	182 (26.11%)	
Hot food intake, n (%)				0.878
Not frequent	1722 (74.03%)	1204 (73.91%)	518 (74.32%)	
Frequent	604 (25.97%)	425 (26.09%)	179 (25.68%)	
Leftover food intake, n (%)				0.618
Not frequent	1809 (77.77%)	1272 (78.08%)	537 (77.04%)	
Frequent	517 (22.23%)	357 (21.92%)	160 (22.96%)	
Mealtime, n (%)				0.246
Unregular	185 (7.95%)	137 (8.41%)	48 (6.89%)	
Regular	2141 (92.05%)	1492 (91.59%)	649 (93.11%)	
Overeat and overdrink, n (%)				0.797
Not frequent	2277 (97.89%)	1596 (97.97%)	681 (97.70%)	
Frequent	49 (2.11%)	33 (2.03%)	16 (2.30%)	
Fresh fruit intake, n (%)				0.392
Not frequent	1547 (66.51%)	1074 (65.93%)	473 (67.86%)	
Frequent	779 (33.49%)	555 (34.07%)	224 (32.14%)	
Fresh vegetables intake, n (%)				0.296
Not frequent	806 (34.65%)	553 (33.95%)	253 (36.30%)	
Frequent	1520 (65.35%)	1076 (66.05%)	444 (63.70%)	
Red meat intake, n (%)				0.285
Not frequent	1435 (61.69%)	993 (60.96%)	442 (63.41%)	
Frequent	891 (38.31%)	636 (39.04%)	255 (36.59%)	
Educational level, n (%)				0.362
illiteracy	239 (10.28%)	174 (10.68%)	65 (9.33%)	
literacy	2087 (89.72%)	1455 (89.32%)	632 (90.67%)	
Digestive tract symptom, n (%)				0.651
No	1133 (48.71%)	788 (48.37%)	345 (49.50%)	
Yes	1193 (51.29%)	841 (51.63%)	352 (50.50%)	
Type of residential area, n (%)				0.544
Rural	2094 (90.03%)	1462 (89.75%)	632 (90.67%)	
Urban	232 (9.97%)	167 (10.25%)	65 (9.33%)	
Occupation, n (%)				0.559
farmer	1887 (81.13%)	1316 (80.79%)	571 (81.92%)	
unfarmer	439 (18.87%)	313 (19.21%)	126 (18.08%)	
Income, n (%)				0.465
High	487 (20.94%)	334 (20.50%)	153 (21.95%)	
Low	1839 (79.06%)	1295 (79.50%)	544 (78.05%)	

(Continued)

TABLE 2 Continued

Variables	Total participants (n = 2326)	Training Cohort (n = 1629)	Validation Cohort (n = 697)	p-value
Marital status, n (%)				0.666
Married	2132 (91.66%)	1490 (91.47%)	642 (92.11%)	
Unmarried	194 (8.34%)	139 (8.53%)	55 (7.89%)	
Current or past manual labour, n (%)				0.082
No	323 (13.89%)	240 (14.73%)	83 (11.91%)	
Yes	2003 (86.11%)	1389 (85.27%)	614 (88.09%)	
BMI (kg/m <sup>2</sup> ), n (%)				0.299
<18.5	228 (9.80%)	167 (10.25%)	61 (8.75%)	
≥18.5	2098 (90.20%)	1462 (89.75%)	636 (91.25%)	
Hp, n (%)				0.775
Negative	2176 (93.55%)	1526 (93.68%)	650 (93.26%)	
Positive	150 (6.45%)	103 (6.32%)	47 (6.74%)	
Water intake (ml/day), n (%)				0.1193
>1000	351 (15.09%)	233 (14.30%)	118 (16.93%)	
≤1000	1975 (84.91%)	1396 (85.70%)	579 (83.07%)	
Chronic disease, n (%)				0.3858
No	1658 (71.28%)	1152 (70.72%)	506 (72.60%)	
Yes	668 (28.72%)	477 (29.28%)	191 (27.40%)	

p values refer to comparison between training cohort and the validation cohort in the univariate analysis.

FHGC, in these two groups are shown in Table 1. The baseline demographic and clinicopathologic characteristics of the development cohort are shown in Table 2, and none of the variables differed significantly between the training and internal validation cohort in all characteristics (all  $p > 0.05$ ), indicating reasonable grouping of participants.

## Prediction factor identification

To identify potentially significant indicators associated with GC development, a LASSO regression algorithm was applied to the training cohort. The age stratification according to published papers (6, 21). The variation characteristics of the coefficient of these variables were shown in Figure 2A. The most appropriate tuning parameter lambda for the LASSO regression was 0.01075537 when the partial likelihood binomial deviance reached its minimum value. After LASSO regression analysis, 10 variables remained significant predictors of neoplastic lesions, including sex, ethnic groups, age, smoked food consumption, alcohol consumption, cigarette smoking, leftover food consumption, education level, BMI, and Hp infection status (Figure 2).

These 10 variables were further selected as candidate variables for multivariable logistic analysis. Six variables, including age, sex, alcohol consumption, cigarette smoking, education level, and Hp

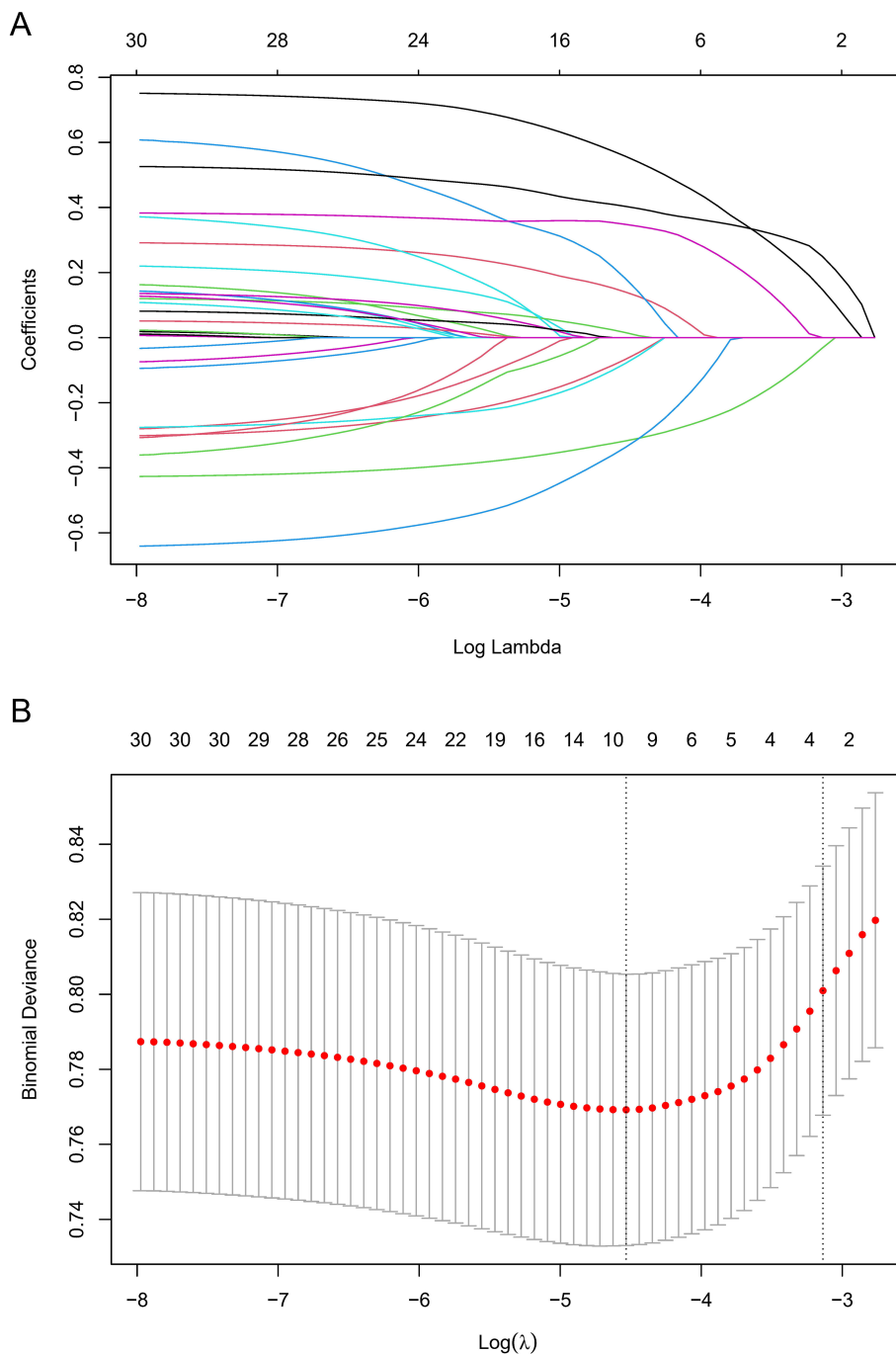
infection status, were independent predictors for the development of neoplastic lesions after multivariable selection (Table 3).

## Nomogram construction

Next, the nomogram used to predict the probabilities of developing neoplastic lesions was formulated using significant independent factors, including age, sex, alcohol consumption, cigarette smoking, education level, and Hp infection status. For an individual participant, the first row shows the points ranging from 0 to 100 received for each variable value which is loaded on each variable axis (row 2-7), and the total score was calculated by summing the individual scores (row 8). The individual probabilities of developing neoplastic lesions were also obtained from the nomogram (Figure 3).

## Assessment of predictive accuracy of the nomogram

The area under the operating characteristic curve (ROC) curve (AUC) values were analyzed to investigate the discrimination of the nomogram, which was 0.701 (95% CI, 0.665-0.738) in the training



**FIGURE 2** Variable selection by LASSO regression model. **(A)** A coefficient profile plot was produced against the log(lambda) sequence. The upper Abscissa is the number of non-zero coefficients in this model, and the ordinate is the coefficient value. **(B)** The optimal penalization coefficient (lambda) in the LASSO model was identified by 10-fold cross-validation via minimum criterion. The left vertical dotted line represents the minimum criterion, and the right line represents the cross-validated error within one standard error of the minimum. The upper Abscissa indicates the number of independent variables that still exist in the model. In the present study, predictor selection was performed according to the minimum criterion. LASSO coefficient profiles of the 30 factors to be screened. The vertical line on the left indicates that the optimal lambda results in the 10 features with non-zero coefficients.

cohort and 0.657 (95% CI, 0.596-0.717) in the internal validation cohort, and 0.699 (95% CI, 0.625-0.773) in the external validation cohort, respectively (Figure 4), indicating that the model has good predictive ability. The calibration curves were also applied to verify the predicted effect of the nomogram, and the calibration plot was

highly consistent between the predicted and observed severity in the training (Figure 5A), the internal validation (Figure 5B), and the external validation cohorts (Figure 5C). The DeLong test results indicated that the predictive performance of the nomogram model was significantly higher than that of the single independent

TABLE 3 Multivariable analysis of risk factors for developing neoplastic lesions.

Variables	Multivariable analysis			
	OR	95% CI lower	95% CI Upper	p-value
<b>Covariates</b>				
<b>Age, years</b>				
<50	1 (ref.)			
50-59	1.46	0.99	2.20	0.061
60-69	2.54	1.68	3.88	0.000
>69	2.96	1.71	5.09	0.000
<b>Sex</b>				
Female	1 (ref.)			
Male	2.14	1.46	3.14	0.000
<b>Ethnic group</b>				
Han	1 (ref.)			
Others	1.32	0.95	1.83	0.091
<b>Smoked food intake</b>				
Not frequent	1 (ref.)			
Frequent	1.31	0.90	1.89	0.157
<b>Alcohol consumption</b>				
No	1 (ref.)			
Yes	1.50	1.04	2.16	0.028
<b>Cigarette smoking</b>				
No	1 (ref.)			
Yes	1.62	1.10	2.40	0.014
<b>Leftover food intake</b>				
Not frequent	1 (ref.)			
Frequent	1.11	0.78	1.57	0.542
<b>Educational level</b>				
illiteracy	1 (ref.)			
literacy	0.55	0.36	0.87	0.008
<b>BMI (kg/m<sup>2</sup>)</b>				
<18.5	1 (ref.)			
≥18.5	0.72	0.47	1.13	0.140
<b>Hp</b>				
Negative	1 (ref.)			
Positive	1.81	1.02	3.11	0.036

OR, odds ratio; CI, Confidence Interval.

predictor in the training cohort, internal validation cohort and external validation cohort, but there was no significant difference between nomogram model and age in the validation cohort (Table 4).

## Clinical application value of nomogram

The decision curve analysis (DCA) analysis was performed to assess the clinical benefit of the model. As shown in Figure 6, the abscissa represents the threshold probability, while the ordinate depicts the net benefit rate after subtracting the pros and cons. The two straight lines in the figure illustrate the two extreme cases. The horizontal line indicates that all participants are considered to be free of neoplastic lesions, thereby indicating a net benefit of 0 in the absence of intervention. The slash denotes the net benefit when all participants are considered to have neoplastic lesions, and all receive the intervention. Our nomogram showed a net benefit across a wide range of threshold probabilities for predicting the risk of developing neoplastic lesions and the nomogram model showed better clinical utility than the predictive value of any single variable, regardless of whether the training, internal validation cohort or external validation cohort was used (Figure 6).

## Discussion

In this study, we established a nomogram model to predict the risk of GC developing using the training cohort and achieved good performance in the training, internal validation, and external validation cohorts. Our finding revealed that advanced age, male sex, alcohol consumption, cigarette smoking, illiteracy, and Hp infection were identified as risk factors that increased the risk of developing GC. Both the calibration curve and the ROC curve indicate that the nomogram has a high accuracy and has a good discrimination. In addition, the DCA shows that the nomogram has good clinical utility. Therefore, this nomogram can serve as a simple, accuracy, and convenient prescreening tool to predict the individual risk of developing GC in China.

A comprehensive understanding of the risk factors for developing GC is essential to identify individuals at high risk who require endoscopic examination. Our multivariable analysis revealed that age, sex, alcohol consumption, cigarette smoking, education level, and Hp infection status were independent predictors of neoplastic lesions (Table 3), which is consistent with findings from previous studies (6, 22–27). It indicates that the predictive factors our selected were reliable. As noted in previous research, the incidence of GC has been shown to increase with age, particularly after 40 years old (6, 22, 23, 28). Consistent with these studies, our results indicated that advanced age was associated with a higher risk of neoplastic lesions developing. Previous studies have shown that men have a higher risk of developing GC than women, suggesting sex also plays an important role in the development of GC (25, 28, 29). This is consistent with the conclusions we have obtained, that male sex has an increased risk for neoplastic lesions. Education level was also significant factor in the prediction model, consistent with the previous meta-analysis (30). In contrast, we did not find significant associations between income level, occupation, and the risk of neoplastic lesions. This is consistent with previous studies (22, 31). Lifestyle factors have been extensively studied in relation to GC risk (32, 33). Unhealthy lifestyles, including cigarette

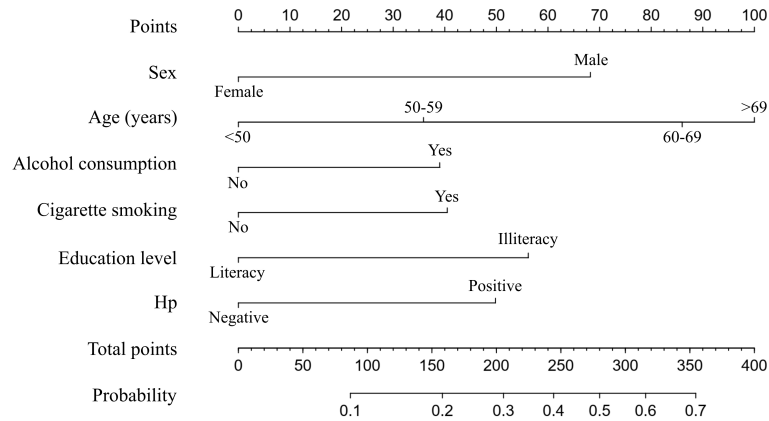


FIGURE 3

A nomogram predicting the probability of developing neoplastic lesions for individual participant. The scores of each variable are added to obtain the total score, and then a vertical line is subtracted from the row of total-points to estimate the probability of developing neoplastic lesions. The neoplastic lesions incidence risk nomogram was developed in the array, with sex, age, alcohol consumption, cigarette smoking, education level, and Hp infection status incorporated.

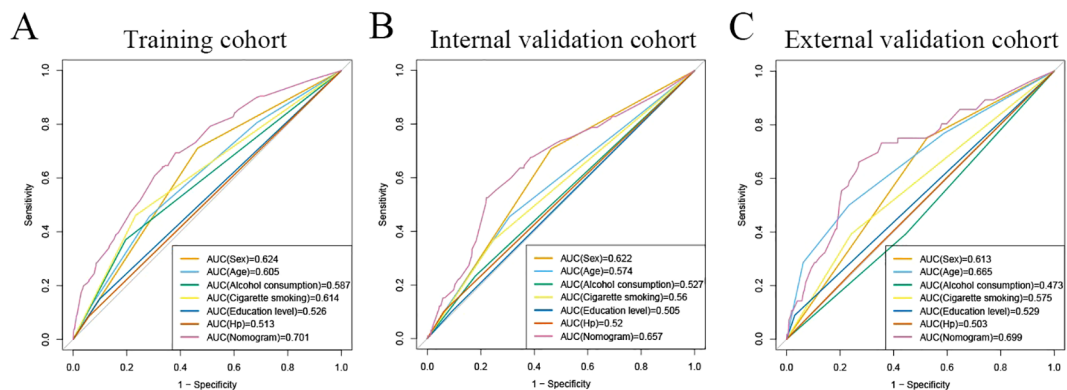


FIGURE 4

ROC curves for predicting risk of developing neoplastic lesions. A receiver operating characteristics (ROC) curve of the nomogram model and single independent predictor in the training cohort (A), internal validation cohort (B), and external validation cohort (C). The model displayed reliable diagnostic performance for prediction of developing neoplastic lesions in both the training cohort (AUC, 0.701), internal validation cohort (AUC, 0.657), and external validation cohort (AUC, 0.699).

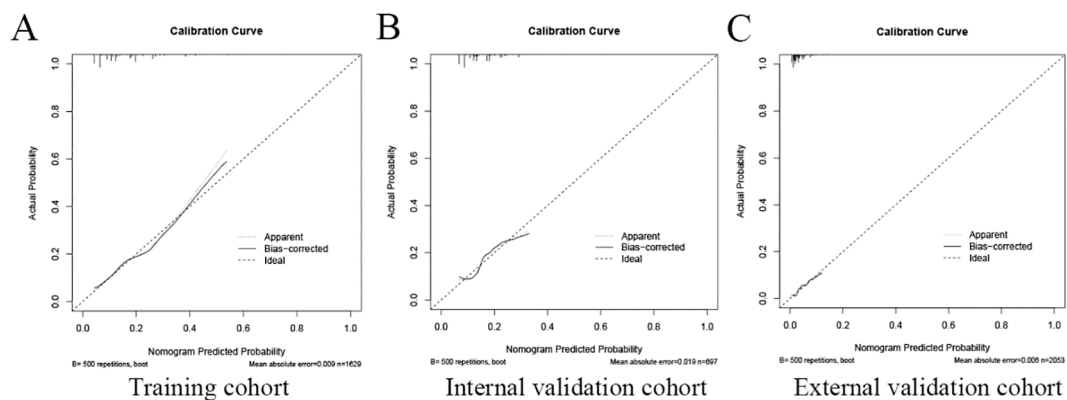


FIGURE 5

The calibration plot of the nomogram in the training and validation cohort. The calibration plot was applied to compare the agreement between actual and predicted probability of developing neoplastic lesions in the training cohort (A), internal validation cohort (B), and external validation cohort (C).

TABLE 4 Delong test between nomogram model and single independent predictor.

Model	p-value		
	Training cohort	Internal validation cohort	External validation cohort
Nomogram vs age	0.0000	0.0036	0.4184
Nomogram vs sex	0.0000	0.0856	0.0006
Nomogram vs alcohol consumption	0.0000	0.0000	0.0000
Nomogram vs cigarette smoking	0.0000	0.0003	0.0007
Nomogram vs educational level	0.0000	0.0000	0.0000
Nomogram vs hp	0.0000	0.0000	0.0000

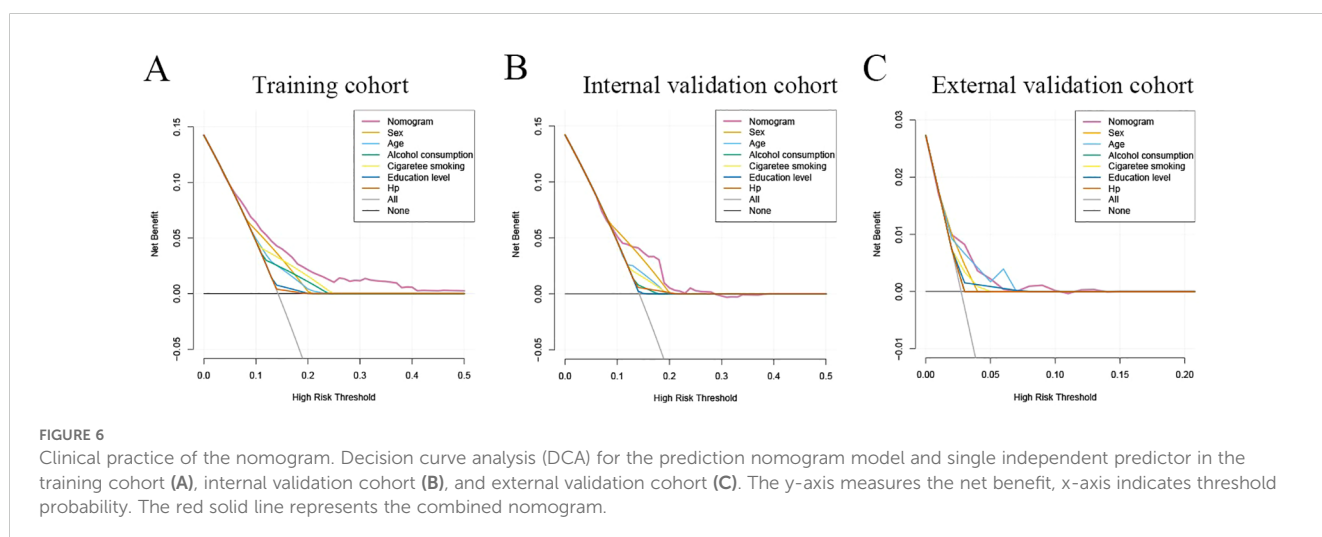
smoking and alcohol consumption, have previously been reported to be associated with a higher risk of GC (26, 34–37). Consistent with these studies, our results indicated that alcohol consumption and cigarette smoking were associated with an increased risk of developing neoplastic lesions. FHGC is well accepted as an important risk factor of GC development (38–40). However, our present study showed that it was weakly correlated with risk of neoplastic lesions (Table 1). This finding can be explained by the low prevalence of FHGC-related neoplastic lesions observed in our study participants. Additionally, the neoplastic lesion group included GIN, which was not considered to be significantly associated with FHGC (41, 42). Conversely, some studies have indicated that there is no positive correlation between FHGC and GC in the Chinese population (22, 43).

Hp infection has been shown to be the most prominent factor for the development of GC (6, 27, 44). More than 90% of GC patients have current or past Hp infection (45, 46). Since the discovery of Hp, numerous studies have confirmed the association between Hp and GC (22, 27, 47). Consistent with previous studies, our results show that Hp infection increases the

risk of developing neoplastic lesions. It has been recommended to eradicate Hp in patients with CAG and IM by the guidelines in Asia and Europe (48), and the eradication of Hp infection is associated with a reduced incidence of GC (49, 50). Therefore, Hp eradication treatment could be used as an important strategy for GC prevention in high-risk populations.

This study has several strengths. Firstly, the utilization of questionnaire data to develop the risk prediction model increases the feasibility of implementing it in real-world screening programs, particularly in areas with inconvenient transport access such as plateau-mountain regions. This approach provides a practical and cost-effective means of identifying individuals at risk of developing GC. Secondly, this newly proposed risk prediction tool has included GIN, which helps to accurately identify individuals with such lesions to further improve their interception and thus potentially reduce the incidence of GC. Third, the performance of this new prediction model was validated in both internally and externally clinical outpatient cohorts with heterogeneous characteristics. Therefore, this model is believed to have a good generalizability and applicability in the Chinese population. Forth, the stability of the Fengqing’s cohort is another strength of the study. The establishment of a stable cohort ensures long-term follow-up of participants, allowing valuable data to be collected over an extended period of time. In addition, the ability to update the cohort in a timely manner allows new information and finding to be incorporated into future analyses, improving the accuracy and relevance of the risk prediction model.

Several limitations of this study need to be noted regarding. First, the number of GC patients in the current cohort we used was relatively small, which may result in the missing of crucial demographic factors about GC development. Further research, particularly in areas with a high incidence of GC, is needed to validate the results and ensure their applicability to different populations. Second, the participants enrolled in our study were from clinical outpatient, which might induce selection bias, even though the study was stringently designed. Future research including the general population is essential to externally validate the results and increase the applicability of the prediction model.



Third, the effect of risk factors for neoplastic lesions might differ between men and women. We performed sex-specific analysis, however, after LASSO regression followed by multivariable logistic analysis, none factor was significant in female participants because of the small number of female cases in neoplastic lesions (Supplementary Table 1). Future studies are warranted to clarify the risk factors of neoplastic lesions in different sex. Fourth, in order for the prediction tool to be widely accepted and used in practice, only a limited number of readily available lifestyle factors were considered. While this increases the feasibility and simplicity of the tool, it may overlook other potential risk factors that may contribute to the development of GC. Further research incorporating a wider range of risk factors, including genetic, and environmental, may provide a more comprehensive understanding of GC risk prediction. In the future, it is expected to combine a wider range of risk factors, such as serological test, plasma metabolite profiling, and genetic factors, with machine learning or deep learning to create high-accuracy prediction models for GC.

## Conclusions

In conclusion, we have developed a low-cost, easy-to-use, and powerful prediction model. The nomogram model developed in this study demonstrated the ability to accurately predict the likelihood of neoplastic lesions in China. This model could serve as a powerful and practical application tool prior to gastroscopy.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving humans were approved by The Ethics Committee of the Fengqing People's Hospital and the Ethics Committee of the Seventh Affiliated Hospital of Sun Yat-Sen University (KY-2021-105-01). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

## Author contributions

ZSH: Conceptualization, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Visualization, Writing – original draft, Writing – review & editing. SC: Conceptualization, Formal Analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. SY: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Writing – review & editing. ZS: Data curation, Funding acquisition, Investigation, Project

administration, Resources, Writing – review & editing. LG: Investigation, Methodology, Visualization, Writing – review & editing. LL: Formal Analysis, Validation, Writing – review & editing. HY: Data curation, Validation, Writing – review & editing. ZJH: Methodology, Validation, Writing – review & editing. BL: Data curation, Validation, Writing – review & editing. XC: Data curation, Resources, Writing – review & editing. YY: Resources, Writing – review & editing, Data curation. ZW: Data curation, Software, Writing – review & editing. HL: Data curation, Funding acquisition, Resources, Supervision, Writing – review & editing. CZ: Conceptualization, Supervision, Writing – review & editing. YH: Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study was funded by the He Yulong Expert Workstation of Yunnan Province (202104AC100001-B03), Science and Technology Special Project of the National Sustainable Development Agenda Innovation Demonstration Zone of Yunnan Province (202104 AC100001-A10), Guangdong Province Key Laboratory of Digestive Cancer Research (2021B1212040006), and Basic and Applied Basic Research Foundation of Guangdong Province (2019A1515110547).

## Acknowledgments

The authors would like to acknowledge the contribution to this study from endoscopy physicians and pathologists at Fengqing People's Hospital.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2024.1351967/full#supplementary-material>

## References

- Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* (2024) 74:229–63. doi: 10.3322/caac.21834
- Han B, Zheng R, Zeng H, Wang S, Sun K, Chen R, et al. Cancer incidence and mortality in China, 2022. *J Natl Cancer Cent.* (2024) 4:47–53. doi: 10.1016/j.jcc.2024.01.006
- Li H, Zhang H, Zhang H, Wang Y, Wang X, Hou H, et al. Survival of gastric cancer in China from 2000 to 2022: A nationwide systematic review of hospital-based studies. *J Glob Health.* (2022) 12:11014. doi: 10.7189/jogh.12.11014
- Fan X, Qin X, Zhang Y, Li Z, Zhou T, Zhang J, et al. Screening for gastric cancer in China: Advances, challenges and visions. *Chin J Cancer Res.* (2021) 33:168–80. doi: 10.21147/j.issn.1000-9604.2021.02.05
- Zeng H, Chen W, Zheng R, Zhang S, Ji JS, Zou X, et al. Changing cancer survival in China during 2003–15: a pooled analysis of 17 population-based cancer registries. *Lancet Glob Health.* (2018) 6:e555–67. doi: 10.1016/S2214-109X(18)30127-X
- Cai Q, Zhu C, Yuan Y, Feng Q, Feng Y, Hao Y, et al. Development and validation of a prediction rule for estimating gastric cancer risk in the Chinese high-risk population: a nationwide multicentre study. *Gut.* (2019) 68:1576–87. doi: 10.1136/gutjnl-2018-317556
- Correa P. Human gastric carcinogenesis: a multistep and multifactorial process—First American Cancer Society Award Lecture on Cancer Epidemiology and Prevention. *Cancer Res.* (1992) 52:6735–40.
- Correa P, Piazuelo MB. The gastric precancerous cascade. *J Dig Dis.* (2012) 13:2–9. doi: 10.1111/j.1751-2980.2011.00550.x
- Shichijo S, Hirata Y, Niikura R, Hayakawa Y, Yamada A, Ushiku T, et al. Histologic intestinal metaplasia and endoscopic atrophy are predictors of gastric cancer development after *Helicobacter pylori* eradication. *Gastrointest Endosc.* (2016) 84:618–24. doi: 10.1016/j.gie.2016.03.791
- Rugge M, Cassaro M, Di Mario F, Leo G, Leandro G, Russo VM, et al. The long term outcome of gastric non-invasive neoplasia. *Gut.* (2003) 52:1111–6. doi: 10.1136/gut.52.8.1111
- Yamada H, Ikegami M, Shimoda T, Takagi N, Maruyama M. Long-term follow-up study of gastric adenoma/dysplasia. *Endoscopy.* (2004) 36:390–6. doi: 10.1055/s-2004-814330
- de Vries AC, van Grieken NC, Looman CW, Casparie MK, de Vries E, Meijer GA, et al. Gastric cancer risk in patients with premalignant gastric lesions: a nationwide cohort study in the Netherlands. *Gastroenterology.* (2008) 134:945–52. doi: 10.1053/j.gastro.2008.01.071
- den Hoed CM, Holster IL, Capelle LG, de Vries AC, den Hartog B, Ter Borg F, et al. Follow-up of premalignant lesions in patients at risk for progression to gastric cancer. *Endoscopy.* (2013) 45:249–56. doi: 10.1055/s-0032-1326379
- Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. *Lancet Oncol.* (2015) 16:e173–180. doi: 10.1016/S1470-2045(14)71116-7
- Rose PG, Java J, Whitney CW, Stehman FB, Lanciano R, Thomas GM, et al. Nomograms predicting progression-free survival, overall survival, and pelvic recurrence in locally advanced cervical cancer developed from an analysis of identifiable prognostic factors in patients from NRG oncology/gynecologic oncology group randomized trials of chemoradiotherapy. *J Clin Oncol.* (2015) 33:2136–42. doi: 10.1200/JCO.2014.57.7122
- Ji R, Zhang Z, Zhang J, Wu Z, Li M, Ye Y, et al. Cohort Profile: A population-based cohort for the study of gastric cancer in northwest area of China. (Wuweii Cohort). *Int J Epidemiol.* (2021) 50:1433–42. doi: 10.1093/ije/dyab083
- Hamashima C. Systematic review G, guideline development group for gastric cancer screening G. Update version of the Japanese guidelines for gastric cancer screening. *Jpn J Clin Oncol.* (2018) 48:673–83. doi: 10.1093/jcco/hyy077
- Hamashima C. Current issues and future perspectives of gastric cancer screening. *World J Gastroenterol.* (2014) 20:13767–74. doi: 10.3748/wjg.v20.i38.13767
- Song C, Zhang R, Wang C, Fu R, Song W, Dou K, et al. Sleep quality and risk of cancer: findings from the English longitudinal study of aging. *Sleep.* (2021), 44. doi: 10.1093/sleep/zsaa192
- Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis. (TRIPOD): the TRIPOD statement. *BMJ.* (2015) 350:g7594. doi: 10.7326/L15-5093-2
- Liu XM, Ma XY, Liu F, Liu ZL, Tang XY, Ji MZ, et al. Gastric cancer screening methods: A comparative study of the Chinese new gastric cancer screening score and kyoto classification of gastritis. *Gastroenterol Res Pract.* (2022) 2022:7639968. doi: 10.1155/2022/7639968
- Chen Z, Zheng Y, Fan P, Li M, Liu W, Yuan H, et al. Risk factors in the development of gastric adenocarcinoma in the general population: A cross-sectional study of the Wuwei Cohort. *Front Microbiol.* (2022) 13:1024155. doi: 10.3389/fmicb.2022.1024155
- Zhou R, Zheng H, Liu M, Liu Z, Guo C, Tian H, et al. Development and validation of a questionnaire-based risk scoring system to identify individuals at high risk for gastric cancer in Chinese populations. *Chin J Cancer Res.* (2021) 33:649–58. doi: 10.21147/j.issn.1000-9604.2021.06.02
- Young E, Philpott H, Singh R. Endoscopic diagnosis and treatment of gastric dysplasia and early cancer: Current evidence and what the future may hold. *World J Gastroenterol.* (2021) 27:5126–51. doi: 10.3748/wjg.v27.i31.5126
- Chen W, Xia C, Zheng R, Zhou M, Lin C, Zeng H, et al. Disparities by province, age, and sex in site-specific cancer burden attributable to 23 potentially modifiable risk factors in China: a comparative risk assessment. *Lancet Glob Health.* (2019) 7:e257–69. doi: 10.1016/S2214-109X(18)30488-1
- Butt J, Varga MG, Wang T, Tsugane S, Shimazu T, Zheng W, et al. Smoking, *Helicobacter pylori* serology, and gastric cancer risk in prospective studies from China, Japan, and Korea. *Cancer Prev Res (Phila).* (2019) 12:667–74. doi: 10.1158/1940-6207.CAPR-19-0238
- Leung WK, Wu MS, Kakugawa Y, Kim JJ, Yeoh KG, Goh KL, et al. Screening for gastric cancer in Asia: current evidence and practice. *Lancet Oncol.* (2008) 9:279–87. doi: 10.1016/S1470-2045(08)70072-X
- Sun D, Lei L, Xia C, Li H, Cao M, He S, et al. Sociodemographic disparities in gastric cancer and the gastric precancerous cascade: A population-based study. *Lancet Reg Health West Pac.* (2022) 23:100437. doi: 10.1016/j.lanwpc.2022.100437
- Pittayanon R, Rerknimitr R, Klaikaew N, Sanpavat A, Chaitongrat S, Mahachai V, et al. The risk of gastric cancer in patients with gastric intestinal metaplasia in 5-year follow-up. *Aliment Pharmacol Ther.* (2017) 46:40–5. doi: 10.1111/apt.2017.46.issue-1
- Rota M, Alicandro G, Pelucchi C, Bonzi R, Bertuccio P, Hu J, et al. Education and gastric cancer risk—An individual participant data meta-analysis in the StoP project consortium. *Int J Cancer.* (2020) 146:671–81. doi: 10.1002/ijc.v146.3
- Uthman OA, Jadidi E, Moradi T. Socioeconomic position and incidence of gastric cancer: a systematic review and meta-analysis. *J Epidemiol Community Health.* (2013) 67:854–60. doi: 10.1136/jech-2012-201108
- van den Brandt PA. The impact of a healthy lifestyle on the risk of esophageal and gastric cancer subtypes. *Eur J Epidemiol.* (2022) 37:931–45. doi: 10.1007/s10654-022-00899-w
- Man J, Ni Y, Yang X, Zhang T, Yuan Z, Chen H, et al. Healthy lifestyle factors, cancer family history, and gastric cancer risk: A population-based case-control study in China. *Front Nutr.* (2021) 8:774530. doi: 10.3389/fnut.2021.774530
- Deng W, Jin L, Zhuo H, Vasilioi V, Zhang Y. Alcohol consumption and risk of stomach cancer: A meta-analysis. *Chem Biol Interact.* (2021) 336:109365. doi: 10.1016/j.cbi.2021.109365
- Li WY, Han Y, Xu HM, Wang ZN, Xu YY, Song YX, et al. Smoking status and subsequent gastric cancer risk in men compared with women: a meta-analysis of prospective observational studies. *BMC Cancer.* (2019) 19:377. doi: 10.1186/s12885-019-5601-9
- Tamura T, Wada K, Tsuji M, Konishi K, Kawachi T, Hori A, et al. Association of alcohol consumption with the risk of stomach cancer in a Japanese population: a prospective cohort study. *Eur J Cancer Prev.* (2018) 27:27–32. doi: 10.1097/CEJ.0000000000000278
- Praud D, Rota M, Pelucchi C, Bertuccio P, Rosso T, Galeone C, et al. Cigarette smoking and gastric cancer in the Stomach Cancer Pooling. (StoP) Project. *Eur J Cancer Prev.* (2018) 27:124–33. doi: 10.1097/CEJ.0000000000000290
- Zhang R, Li H, Li N, Shi JF, Li J, Chen HD, et al. Risk factors for gastric cancer: a large-scale, population-based case-control study. *Chin Med J (Engl).* (2021) 134:1952–8. doi: 10.1097/CM9.0000000000001652
- Choi JJ, Kim CG, Lee JY, Kim YI, Kook MC, Park B, et al. Family history of gastric cancer and *Helicobacter pylori* treatment. *N Engl J Med.* (2020) 382:427–36. doi: 10.1056/NEJMoa1909666
- Choi YJ, Kim N. Gastric cancer and family history. *Korean J Intern Med.* (2016) 31:1042–53. doi: 10.3904/kjim.2016.147
- Murphy JD, Epplein M, Lin FC, Troester MA, Nichols HB, Butt J, et al. Discrimination between precancerous gastric lesions and gastritis using a gastric cancer risk stratification model. *Asian Pac J Cancer Prev.* (2023) 24:935–43. doi: 10.31557/APJCP.2023.24.3.935
- Liu Y, Cai Y, Chen S, Gou Y, Wang Q, Zhang M, et al. Analysis of risk factors of gastric low-grade intraepithelial neoplasia in asymptomatic subjects undergoing physical examination. *Gastroenterol Res Pract.* (2020) 2020:7907195. doi: 10.1155/2020/7907195
- Huang D, Song M, Abe SK, Rahman MS, Islam MR, Saito E, et al. Family history and gastric cancer incidence and mortality in Asia: a pooled analysis of more than half a million participants. *Gastric Cancer.* (2024) 27:701–13. doi: 10.1007/s10120-024-01499-1
- Nie Y, Wu K, Yu J, Liang Q, Cai X, Shang Y, et al. A global burden of gastric cancer: the major impact of China. *Expert Rev Gastroenterol Hepatol.* (2017) 11:651–61. doi: 10.1080/17474124.2017.1312342



45. Kato S, Matsukura N, Tsukada K, Matsuda N, Mizoshita T, Tsukamoto T, et al. Helicobacter pylori infection-negative gastric cancer in Japanese hospital patients: incidence and pathological characteristics. *Cancer Sci.* (2007) 98:790–4. doi: 10.1111/j.1349-7006.2007.00478.x
46. Yoon H, Kim N, Lee HS, Shin CM, Park YS, Lee DH, et al. Helicobacter pylori-negative gastric cancer in South Korea: incidence and clinicopathologic characteristics. *Helicobacter.* (2011) 16:382–8. doi: 10.1111/j.1523-5378.2011.00859.x
47. Karimi P, Islami F, Anandasabapathy S, Freedman ND, Kamangar F. Gastric cancer: descriptive epidemiology, risk factors, screening, and prevention. *Cancer Epidemiol Biomarkers Prev.* (2014) 23:700–13. doi: 10.1158/1055-9965.EPI-13-1057
48. Yoon H, Kim N. Diagnosis and management of high risk group for gastric cancer. *Gut Liver.* (2015) 9:5–17. doi: 10.5009/gnl14118
49. Yan L, Chen Y, Chen F, Tao T, Hu Z, Wang J, et al. Effect of helicobacter pylori eradication on gastric cancer prevention: updated report from a randomized controlled trial with 26.5 years of follow-up. *Gastroenterology.* (2022) 163:154–162 e153. doi: 10.1053/j.gastro.2022.03.039
50. Jung YS, Tran MTX, Song H, Park B, Moon CM. Association between Age at Helicobacter pylori Eradication and the Risk of Gastric Cancer Stratified by Family History of Gastric Cancer: A Nationwide Population-Based Study. *Cancers (Basel).* (2023) 15, 1604. doi: 10.3390/cancers15051604