scientific reports

Check for updates

OPEN

Derivation and validation of a nomogram incorporating modifiable lifestyle factors to predict development of colorectal adenomas after negative index colonoscopy

Mingqian Yu[1,8], Yiben Ouyang[1,8], Zhen Yuan[1,2,8], Shuyuan Wang[1,2,8], Wenwen Pang[5], Suying Yan[2,3], Xinyu Liu[4], Wanting Wang[2,3], Ben Yi[2,3], Qiurong Han[2,3], Yao Yao[2,3], Yanfei Liu[2,3], Jiachun Song[1], Tianhao Chu[2,3], Zhiqiang Feng[2,3], Qinghuai Zhang[2,6,7], Xipeng Zhang[2,6,7]✉ & Chunze Zhang[1,2,6,7]✉

This retrospective cohort study aimed to identify baseline patient characteristics involving modifiable lifestyle factors that are associated with the development of colorectal adenomas, and establish and validate a nomogram for risk predictions among high-risk populations with negative index colonoscopy. A total of 83,076 participants who underwent an index colonoscopy at the Tianjin Union Medical Center between 2004 and 2019 were collected. According to meticulous inclusion and exclusion criteria, 249 subjects were enrolled and categorized into the primary and validation cohorts. Based on the primary cohort, we utilized the LASSO-Cox regression and the univariate/multivariate Cox proportional hazards (Cox-PH) regression parallelly to select variables, and incorporated selected variables into two nomogram models established using the multivariate Cox-PH regression. Comparison of the Akaike information criterion and the area under the receiver operating characteristic curve of the two models demonstrated that the nomogram model constituted by four covariates retained by the LASSO-Cox regression, including baseline age, body mass index, physical activity and family history of colorectal cancer (CRC) in first-degree relatives, performed better at predicting adenoma-free survival probabilities. Further validation including the concordance index, calibration plots, decision curve analysis and Kaplan–Meier survival curves also revealed good predictive accuracy, discriminating ability, clinical utility and risk stratification capacity of the nomogram model. Our nomogram will assist high-risk individuals with negative index colonoscopy to prevent colorectal adenoma occurrence and CRC morbidity with improved cost-effectiveness.

Colorectal cancer (CRC) ranks third in terms of incidence among cancers and remains the second leading cause of cancer-related mortality globally, with more than 1.06 million new CRC cases and 515,000 deaths estimated to occur in men, and more than 0.86 million new CRC cases and 419,000 deaths estimated to occur in women, based on the GLOBOCAN 2020 estimates[1]. In the past few decades, China has undergone a convergence toward the common cancer profiles in developed countries, characterized by the high incidence and mortality of CRC,

[1]School of Medicine, Nankai University, Tianjin 300071, China. [2]Department of Colorectal Surgery, Tianjin Union Medical Center, Tianjin 300121, China. [3]School of Integrative Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin 301617, China. [4]Tianjin Medical University, Tianjin 300041, China. [5]Department of Clinical Laboratory, Tianjin Union Medical Center, Tianjin 300121, China. [6]The Institute of Translational Medicine, Tianjin Union Medical Center, Tianjin 300121, China. [7]Tianjin Institute of Coloproctology, Tianjin, China. [8]These authors contributed equally: Mingqian Yu, Yiben Ouyang, Zhen Yuan and Shuyuan Wang. ✉email: xipengzhangtj@163.com; chunze.zhang@nankai.edu.cn

in parallel with an escalating adult population size, swift population aging, socioeconomic developments and especially, accumulative exposure to lifestyle risk factors that prevail in western countries[1–3].

The development of most sporadic CRC follows the adenoma-carcinoma sequence[4], and colorectal adenomas are neoplastic lesions of the large bowel that are widely recognized as precursors for the vast majority of CRC cases[5]. Earlier detection of CRC can improve the prognosis prominently since the 5-year relative survival rate is up to 90% when the lesions remain localized[6]. Interrupting this adenoma-carcinoma sequence with population-based screening programs is thus warranted to boost timely detection and early prevention of CRC.

A series of screening modalities for CRC are currently available, including fecal immunochemical test (FIT), colonoscopy and flexible sigmoidoscopy, etc. Among them, colonoscopy is endorsed as the gold standard[7,8]. Considering the large population and the overwhelming constraints of healthcare resources, it is impractical for China to provide a "one-size-fit-all" screening service across each province and county. Nowadays, most CRC screening programs in China adopt a two-step screening strategy, which first identifies individuals with a positive result of either high-risk factor questionnaire (HRFQ) or FIT as high-risk individuals, and subsequently recommends them to undergo colonoscopy using repeat screening[9]. However, the effectiveness of current risk stratification systems has been limited by incremental healthcare expenditures[10], insufficient endoscopic capacity[11] and relatively low participation rates[7,8]. In the era of precision medicine, the exploitation of state-of-the-art techniques and suitable risk-adapted screening strategies based on sophisticated risk prediction models which incorporate multisectoral risk factors is of paramount significance for implementing individualized intervention[8,12].

Because of the general absence of pertinent symptoms, the frequency of colorectal adenomas has commonly been calculated as prevalence rates rather than incidence rates in different populations. The prevalence rates are usually computed based on each participant with a positive index colonoscopy. Nevertheless, actual adenoma incidence rates can be described only by follow-up colonoscopies in each individual whose initial colonoscopy was negative[13,14]. Swelling ranks of studies have reiterated that various modifiable lifestyle factors including smoking status stas[15,16], alcohol consumption[17,18], and physical activity[19,20] are independently and significantly correlated with the presence and development of colorectal adenomas. However, current risk scoring models in predicting risks of developing colorectal neoplasms predominantly rely on demographic and clinicopathological features and often ignore some well-documented lifestyle factors. These unmeasured confounders may bias the observed correlation and the weighing of collected risk factors, thus jeopardizing the performance and utility of models[21].

In the present study, we embarked on a retrospective cohort study based on the first surveillance colonoscopies that high-risk individuals underwent after negative index colonoscopy. We incorporated various modifiable lifestyle patterns with demographic and clinical characteristics of subjects to establish a nomogram, which could reduce statistical predictive models into an intuitive numerical estimate of the probability of certain event[22], for predicting risks of developing colorectal adenomas within different time periods for each subject. In a gesture to facilitate risk stratification, our model will provide ponderable reference for the primary and secondary prevention of colorectal adenomas and CRC.

Methods

Study design and population

Screening of high-risk individuals

In conformity with the two-step screening strategy, each participant was required to first complete the HRFQ and then undergo FIT. The HRFQ was mainly comprised of four questions: (1) with or without a personal history of cancer; (2) with or without a personal history of colorectal adenomas; (3) with or without a family history of CRC in first-degree relatives (FDR); (4) with or without a personal history of two or more items of the following conditions: chronic constipation, chronic diarrhea, mucous bloody stools, adverse life events, chronic appendicitis or appendectomy, and chronic cholecystitis or cholecystectomy. An affirmative answer to any of the four questions was defined as a positive result of HRFQ. Participants with a positive result of either HRFQ or FIT were classified into the high-risk group while the remaining were categorized into the average-risk group. Screening physicians would intensively recommend high-risk participants to undergo colonoscopy examination. According to the findings of index colonoscopy, high-risk individuals were subsequently recommended to undergo surveillance colonoscopies in conformity with authoritative guidelines and were followed up by trained investigators.

Inclusion of the study population

Based on electronic medical records, the first-recorded colonoscopy that each high-risk individual underwent within our study period was considered as the index colonoscopy. The index colonoscopy was designated as a positive colonoscopy if there existed any colorectal adenoma, whereas a negative index colonoscopy denoted that there existed no abnormalities.

A total of 83,076 participants who underwent an index colonoscopy with available records at the Tianjin Union Medical Center between March 16, 2004, and December 31, 2019, were collected. Individuals aged under 40 years or over 75 years were considered ineligible (n = 2655). Individuals with a personal history of cancer (n = 1506) or colorectal adenomas (n = 3562) before the index colonoscopy were also rejected. Moreover, 2410 subjects whose index colonoscopies were considered unqualified were excluded. Among the remaining 72,943 participants, subjects who were diagnosed with other colorectal diseases including non-adenomatous colorectal polyps, CRC, inflammatory bowel disease (Crohn's disease or ulcerative colitis), familial adenomatous polyposis and P-J syndrome, etc. at the index colonoscopy were excluded (n = 6916). 33,752 subjects diagnosed with colorectal adenomas at the index colonoscopy were also dismissed. Among the remaining 32,275 subjects who had negative index colonoscopy, those who underwent at least one surveillance colonoscopy within the study period with the first surveillance colonoscopy being conducted later than 6 months after the index colonoscopy

were enrolled into this retrospective cohort study (n = 571). As discovered by the first surveillance colonoscopy and confirmed by pathological reports, 55 subjects developed aforementioned other colorectal diseases, and 290 subjects had newly developed colorectal adenomas, while 226 subjects had no abnormalities. Each subject whose first surveillance colonoscopy was considered unqualified was eliminated (n = 7 for the negative group and n = 9 for the positive group). Individuals who provided unreliable or incomplete baseline information related to any of the studied factors were also excluded (n = 65 for the negative group and n = 186 for the positive group).

Ultimately, 249 qualified subjects consisted of 95 persons in the occurrence group and 154 persons in the non-occurrence group were categorized into the primary cohort (n = 174) and the validation cohort (n = 75) according to the chronological order of dates of index colonoscopy at a ratio of 7:3.

Colonoscopy examination and pathological diagnosis

All colonoscopies were performed by a group of skilled endoscopists who were certified by the medical facility's Endoscopy Committee and had at least five years of experience. Only complete colonoscopies with successful cecal intubation, photo documentation of caecal landmarks, adequate bowel preparation and a withdrawal time > 6 min were considered as qualified colonoscopies and included for analysis according to up-to-date clinical guidelines[23–26]. The endoscopists removed all polyps detected, and forwarded them to the pathological laboratory, where the histology of polyps was further confirmed by pathologists. Pathological diagnosis was made by two experienced pathologists separately, and only concordant results were adopted.

Outcomes and definitions

The outcome in this study was adenoma-free survival (AFS). Similar to previous literature, we excluded subjects with their first surveillance colonoscopy being conducted within six months after negative index colonoscopy, because any adenoma which was detected at subsequent colonoscopies performed within six months after the index colonoscopy was perceived as a missing adenoma that had already existed at the index colonoscopy[27,28].

Colorectal polyps were classified as neoplastic adenomatous polyps (tubular, tubular villous or villous) or non-neoplastic polyps (hyperplastic, hamartomatous or inflammatory) in accordance with standard criteria[29]. A polyp with a mixed histology (e.g., both hyperplastic and adenomatous components) was considered as an adenoma. BMI was calculated as weight in kilograms divided by height in meters squared. Participants were categorized into never-smokers and smokers according to their smoking status, with the former including subjects who never smoke, and the latter comprised of subjects who smoke currently (smoking at least one pack of cigarettes per week over the last year) or in the past. Participants were also classified into never-drinkers and drinkers in terms of alcohol consumption, with the former including subjects who never drank, and the latter comprised of subjects who drank occasionally, frequently or even daily over the last year. Physical activity was sorted into regular activity and physical inactivity estimated by the frequency of exercise. The former denotes at least 30 min of exercise more than once weekly over the last year, and the latter is defined otherwise. Chronic constipation refers to constipation lasting for more than two months per year in the last two years. Chronic diarrhea refers to diarrhea lasting for more than one week each time, with a cumulative course of more than three months per year in the last two years.

Formulation and validation of the nomogram

Ten variables encompassing demographic and clinical characteristics of subjects as well as various modifiable lifestyle patterns were chosen as candidate predictors for colorectal adenoma occurrence anchored in previous literature, including baseline age[7,16,30,31], BMI[32], gender[30,33], smoking status[15,16], alcohol consumption[17,18], physical activity[19,20], family history of CRC in FDR[7,34], history of chronic constipation[35], history of chronic diarrhea[31,36], and history of chronic appendicitis or appendectomy[36]. Two methods were deployed to screen out putative risk predictors from the ten variables for construction of nomogram models based on the primary cohort. Firstly, using the LASSO-Cox method (Method-1), variables with non-zero coefficients retained by the LASSO regression were entered into the multivariate Cox proportional hazards (Cox-PH) regression. Variables with $p < 0.05$ were considered as independent predictors and incorporated into a nomogram model (Nomogram-1) constructed using the multivariate Cox-PH regression. Secondly, using the univariate/multivariate Cox-PH regression (Method-2), variables with $p < 0.05$ as revealed by the univariate Cox-PH regression were entered into the multivariate Cox-PH regression analysis. The independent predictors determined by the multivariate analysis were absorbed into Nomogram-2 formulated using the multivariate Cox-PH regression. Performance of the two nomograms was compared by calculating the Akaike information criterion (AIC) and the area under the receiver operating characteristic (ROC) curve (AUC) values. Anchored in the concept of entropy, AIC is a measure of the goodness of fit of regression models. A lower AIC value denotes a better model fit[37]. Meanwhile, a higher AUC value signifies a better discriminative power of the model[38]. Ultimately, the nomogram with lower AIC and higher AUC values was selected to predict risks of colorectal adenoma occurrence. The proportional hazards (PH) assumption, which is the underlying premise for the Cox-PH model, was tested by Schoenfeld residual plots and deviance residual before performing the univariate or multivariate Cox-PH regression analysis. Additionally, multicollinearity among predictors in the multivariate Cox-PH regression analysis was assessed using variation inflation factor (VIF), and only predictors with VIF values < 5 were entered[39].

To evaluate the performance of the nomogram, we calculated the risk score for each subject and computed the optimal cut-off value of the nomogram in the primary cohort and the validation cohort respectively. The utility of the model was judged by the sensitivity (SE), specificity (SP), positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (PLR) and negative likelihood ratio (NLR) for subject stratification into the occurrence group and the non-occurrence group based on the optimal cut-off values in both cohorts. The accuracy of the selected nomogram was appraised via two approaches. Firstly, the predictive accuracy for

individual outcomes (discrimination) was determined using the concordance index (C-index) values[38]. Secondly, the precision of point estimates of the survival function (calibration) was assessed by comparing the predicted and actual probability of outcomes. Bootstraps with 1000 resamples were employed for these activities. Moreover, clinical utility of the nomogram was judged by the decision curve analysis (DCA). To verify the risk stratification ability of the nomogram, subjects were classified into the high-risk group and the low-risk group according to the optimal cut-off value of risk scores in each cohort. The Kaplan–Meier (KM) survival curves with log-rank tests were utilized to assess whether the survival outcomes were significantly different between the two groups.

Statistical analysis

The Kolmogorov–Smirnov test was employed to examine the distribution of continuous variables. Continuous variables with normal distribution were presented as mean with standard error and compared using the Student's *t*-test. Continuous variables with non-normal distribution were presented as median with interquartile ranges (IQRs) and compared using the Wilcoxon–Mann–Whitney test. Categorical variables were exhibited as whole numbers with proportions and compared using the Pearson Chi-square test or continuity correction Chi-square test. Statistical analysis was performed with SPSS software (version 19.0, SPSS Inc., Chicago, IL, USA) and R software (version 4.0.2, http://www.r-project.org). $P < 0.05$ was considered as statistically significant.

Ethics approval and consent to participate

This study was conducted following the Declaration of Helsinki and approved by the Ethics Committee of Tianjin Union Medical Center. All participants provided their written informed consent and their privacy and identities were fully protected in the manuscript.

Results

Baseline characteristics of the study population

The procedure of subject inclusion is shown in Fig. 1. The process of variable selection, nomogram construction and performance validation is depicted in Fig. 2. According to meticulous inclusion and exclusion criteria, 249 eligible participants were enrolled and divided into two independent cohorts (Fig. 1). Participants who underwent an index colonoscopy between September 29, 2004, and May 6, 2016, were included in the primary cohort (n = 174) to construct the nomogram, and those who underwent an index colonoscopy between May 11, 2016, and April 10, 2019, were treated as the validation cohort (n = 75) to appraise model performance. At a median

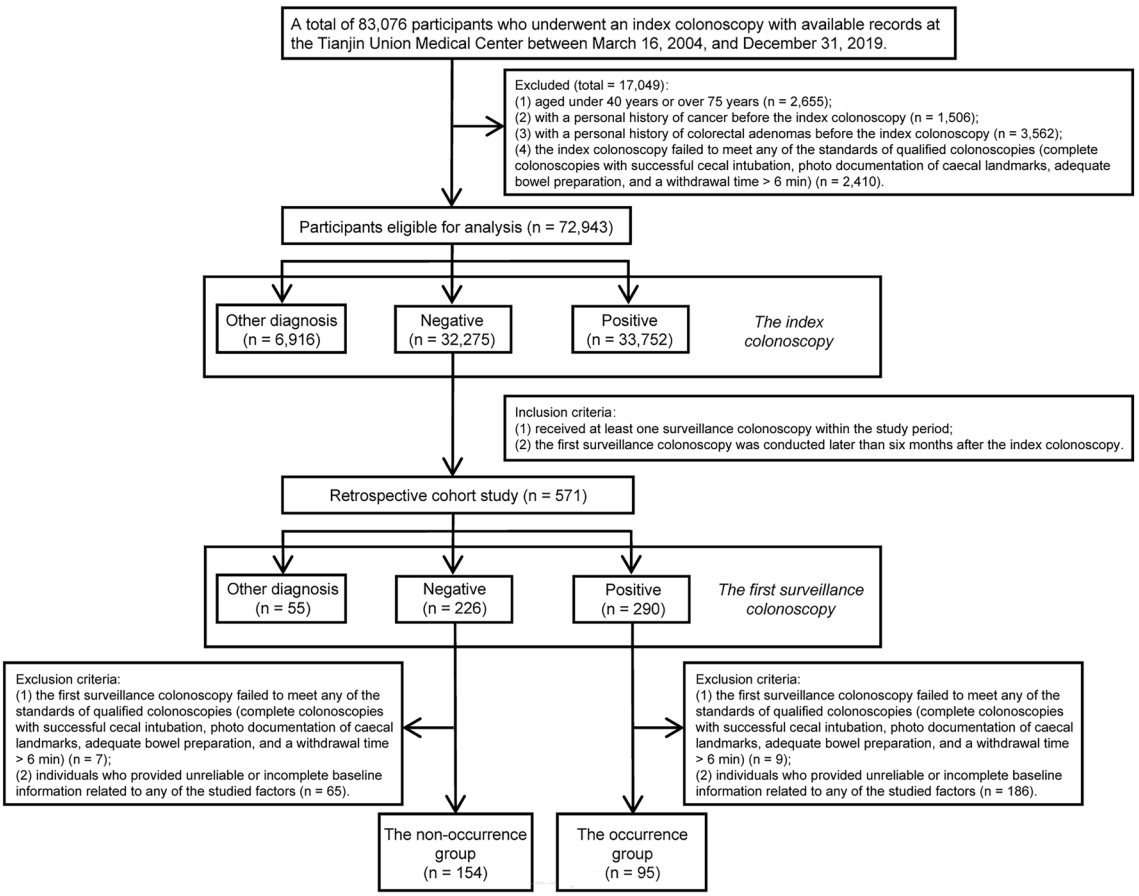


Figure 1. Flow diagram depicting the procedure of subject inclusion.

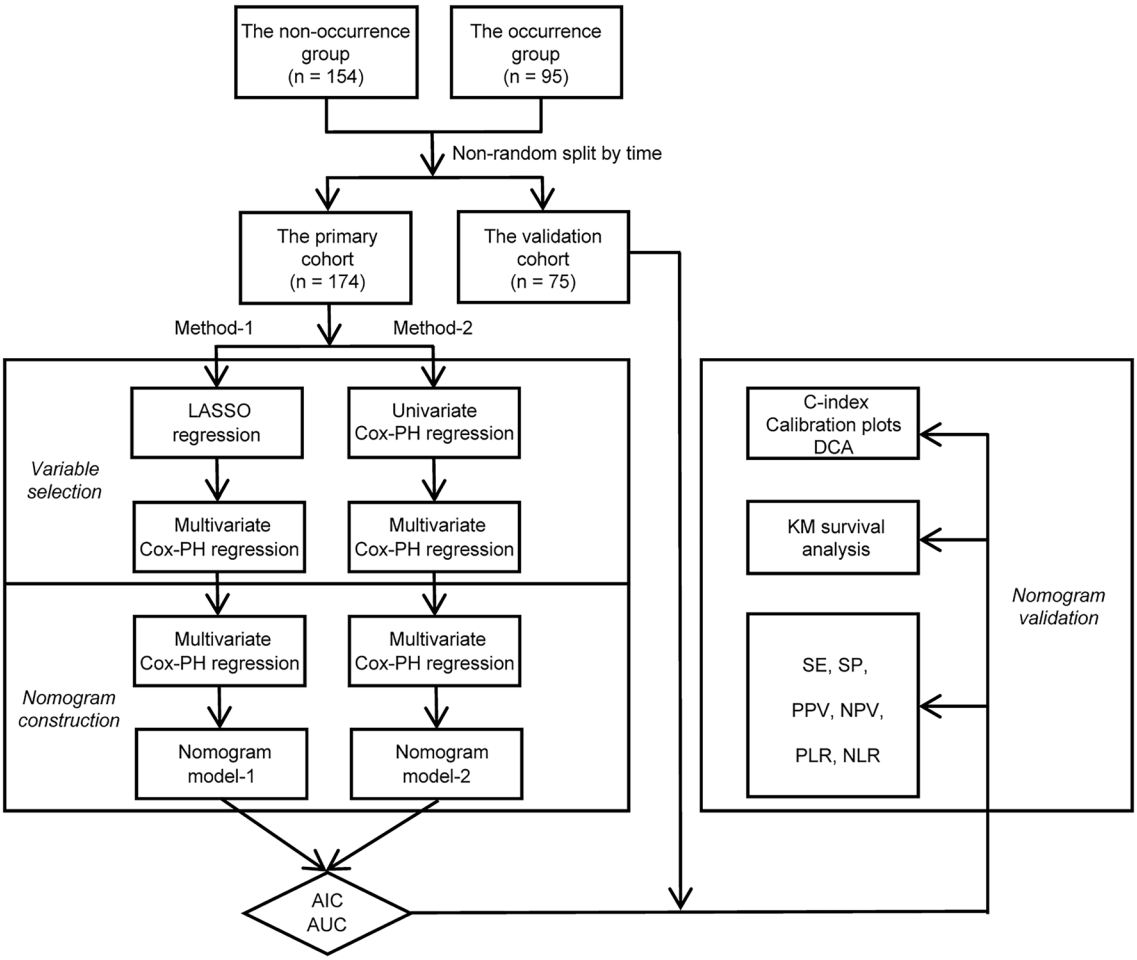


Figure 2. Flow diagram depicting the process of variable selection, nomogram construction and performance validation. LASSO: least absolute shrinkage and selection operator; Cox-PH: Cox proportional hazards; AIC: Akaike information criterion; AUC: area under the curve; C-index: concordance index; DCA: decision curve analysis; KM: Kaplan–Meier; SE: sensitivity; SP: specificity; PPV: positive predictive value; NPV: negative predictive value; PLR: positive likelihood ratio; NLR: negative likelihood ratio.

(range) follow-up time of 25.40 months (6.07–144.47), 38.15% (95 of 249) of the subjects developed colorectal adenomas. The 12-month, 24-month, and 36-month AFS percentages were 93.57%, 81.93% and 75.90%, respectively.

The demographic and clinical characteristics of subjects in the primary cohort and the validation cohort were demonstrated in Table 1. In the primary cohort, 58.05% (n = 101) of the subjects were female, with a median (IQR) age of 61.134 (58.579, 64.411) years, and a median (IQR) BMI of 23.931 (22.656, 25.952) kg/m^2. Moreover, 14.94% (n = 26) of them were current or past smokers, and 14.94% (n = 26) of them drank occasionally, frequently or even daily. Additionally, 43.68% (n = 76) of them shared the common feature of physical inactivity, and 12.07% (n = 21) of them had a family history of CRC in FDR. The percentages of individuals who had a personal history of chronic constipation, chronic diarrhea, and chronic appendicitis or appendectomy were 29.89% (n = 52), 28.74% (n = 50) and 10.34% (n = 18), respectively. Notwithstanding a temporal disconnect, the characteristics of subjects in the validation cohort were similar to those of subjects in the primary cohort. Only age ($p < 0.0001$) and physical activity ($p = 0.003$) were significantly different between the two cohorts. The characteristics of subjects in the occurrence group and the non-occurrence group within each cohort were exhibited in Supplementary Table S1.

Identification of predictive factors and construction of nomogram

Method-1

In the LASSO regression analysis based on tenfold cross-validation (Fig. 3A,B), a λ value of 0.013, which corresponded with a log(λ) of −4.360, was chosen according to the minimum criteria. Eight variables with non-zero coefficients were retained, including baseline age, BMI, physical activity, history of chronic constipation, gender, smoking status, alcohol consumption, and family history of CRC in FDR. As shown in Fig. 3C, the multivariate Cox-PH regression analysis disclosed that among these eight variables, age (hazard ratio [HR] = 1.131, 95% confidence interval [CI] = 1.066–1.200, $p < 0.0001$), BMI (HR = 1.124, 95% CI = 1.020–1.238, $p = 0.018$), physical activity (HR = 0.526, 95% CI = 0.303–0.914, $p = 0.023$) and family history of CRC in FDR (HR = 3.335, 95% CI = 1.394–7.977, $p = 0.007$) were independent predictors for the development of colorectal adenomas, while

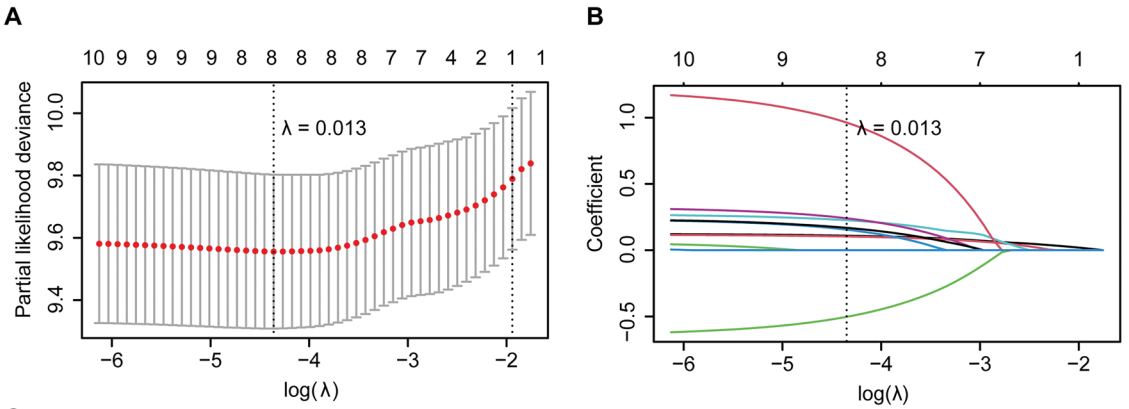
Variables	The primary cohort (n = 174)	The validation cohort (n = 75)	*p*-Value
Age (median (IQRs))	61.134 (58.579, 64.411)	55.586 (51.737, 60.864)	< 0.0001
BMI (median (IQRs))	23.931 (22.656, 25.952)	23.225 (22.231, 25.064)	0.094
Gender (n (%))			0.765
Female	101 (58.05)	42 (56.00)	
Male	73 (41.95)	33 (44.00)	
Smoking status (n (%))			0.324
Never	148 (85.06)	60 (80.00)	
Current or past	26 (14.94)	15 (20.00)	
Alcohol consumption (n (%))			0.324
Never	148 (85.06)	60 (80.00)	
Occasional, frequent or even daily	26 (14.94)	15 (20.00)	
Physical activity (n (%))			0.003
Physical inactivity	76 (43.68)	48 (64.00)	
Regular activity	98 (56.32)	27 (36.00)	
Family history of CRC in FDR (n (%))			0.752
No	153 (87.93)	67 (89.33)	
Yes	21 (12.07)	8 (10.67)	
History of chronic constipation (n (%))			0.465
No	122 (70.11)	56 (74.67)	
Yes	52 (29.89)	19 (25.33)	
History of chronic diarrhea (n (%))			0.924
No	124 (71.26)	53 (70.67)	
Yes	50 (28.74)	22 (29.33)	
History of chronic appendicitis or appendectomy (n (%))			0.098
No	156 (89.66)	72 (96.00)	
Yes	18 (10.34)	3 (4.00)	

Table 1. Baseline information of demographic and clinical characteristics of subjects in the primary cohort and the validation cohort. Continuous variables were compared using the Wilcoxon–Mann–Whitney test. Categorical variables were compared using the Pearson Chi-square test. All tests were two-sided. *IQR* interquartile range, *BMI* body mass index, *CRC* colorectal cancer, *FDR* first-degree relative.

other parameters demonstrated no significant association ($p > 0.05$). No VIF ≥ 5 was detected (Fig. 3C), which excluded the possibility of multicollinearity among variables and justified our use of the multivariate Cox-PH regression. As demonstrated in Fig. 4A–H, the results of Schoenfeld's individual and global test indicated no violation of the PH assumption ($p > 0.05$), proving that none of the covariates were time-varying. The results of deviance residual also uncovered that none of the individual observations were highly influential (Fig. 4I–P). Subsequently, we incorporated these four variables into Nomogram-1 (Fig. 5A) formulated using the multivariate Cox-PH regression, the result of which was demonstrated in Table 2. No VIF ≥ 5 (Table 2) or breach of the PH assumption ($p > 0.05$, Supplementary Figure S1A–D) was detected, and there existed no greatly influential individual observations (Supplementary Figure S1E–H). The AIC, 24-month AUC, 30-month AUC and 36-month AUC values were calculated as 502.451, 0.657, 0.638, and 0.642 respectively.

Method-2

The univariate Cox-PH regression analysis elucidated that older age (HR = 1.099, 95% CI = 1.048–1.153, $p = 0.001$), higher BMI (HR = 1.147, 95% CI = 1.038–1.269, $p = 0.007$) and male gender (HR = 1.743, 95% CI = 1.073–2.832, $p = 0.025$) were significantly associated with an elevated risk of developing colorectal adenomas (Table 3). The results of Schoenfeld's tests and deviance residual were shown in Supplementary Figure S2. The multivariate Cox-PH regression analysis further disclosed that among the three variables, only age (HR = 1.102, 95% CI = 1.047–1.161, $p < 0.0001$, Table 3) and BMI (HR = 1.133, 95% CI = 1.031–1.246, $p = 0.010$) were independent predictors for the development of colorectal adenomas, while other parameters demonstrated no significant association ($p > 0.05$). No VIF ≥ 5 was detected (Table 3), and the results of Schoenfeld's test and deviance residual suggested no breach of the PH assumption ($p > 0.05$, Supplementary Figure S3A–C) or greatly influential individual observations (Supplementary Figure S3D–F). Subsequently, we incorporated these two variables into Nomogram-2 (Fig. 5B) formulated using multivariate Cox-PH regression, the result of which was demonstrated in Table 2. No VIF ≥ 5 (Table 2) or breach of the PH assumption ($p > 0.05$, Supplementary Figure S4A,B) was detected, and there existed no greatly influential individual observations (Supplementary Figure S4C,D). The AIC, 24-month AUC, 30-month AUC and 36-month AUC values were calculated as 506.608, 0.590, 0.603 and 0.605 respectively.



C

Variables	Patients	HR (95% CI)	*p*-Value	VIF
Age	N = 174	1.131 (1.066 − 1.200)	< 0.0001	1.164
BMI	N = 174	1.124 (1.020 − 1.238)	0.018	1.039
Physical activity			0.023	1.084
Physical inactivity	N = 76	Reference		
Regular activity	N = 98	0.526 (0.303 − 0.914)		
Chronic constipation			0.432	1.064
No	N = 122	Reference		
Yes	N = 52	1.267 (0.702 − 2.288)		
Gender			0.392	1.396
Female	N = 101	Reference		
Male	N = 73	1.310 (0.706 − 2.429)		
Smoking status			0.368	1.305
Never	N = 148	Reference		
Current or past	N = 26	1.387 (0.680 − 2.829)		
Alcohol consumption			0.515	1.216
Never	N = 148	Reference		
Occasional, frequent or even daily	N = 26	1.264 (0.625 − 2.556)		
Family history of CRC in FDR			0.007	1.090
No	N = 153	Reference		
Yes	N = 21	3.335 (1.394 − 7.977)		

Figure 3. Variable selection through the LASSO-Cox regression for predicting risks of colorectal adenoma occurrence based on the primary cohort. (**A**) Tuning parameter (λ) selection of deviance in the LASSO regression analysis. The red dot denotes the CVM, and the gray line stands for the SE of CVM. (**B**) LASSO coefficient profiles of ten candidate variables. Each curve in different color signifies the trajectory of the coefficient of each variable. (**C**) A forest plot exhibiting results of the multivariate Cox-PH regression analysis. CVM: mean cross-validated error; SE: standard error; HR: hazard ratio; CI: confidence interval; VIF: variance inflation factor; BMI: body mass index; CRC: colorectal cancer; FDR: first-degree relatives.

Consequently, we chose Nomogram-1, which harbored both lower AIC and higher AUC values than Nomogram-2, for predicting risks of colorectal adenoma occurrence. Within Nomogram-1, points are assigned by drawing a line upward from the corresponding values to the "Points" line. The sum of these four points, plotted downward on the "Total Points" line, corresponds with predictions of 12-month, 24-month, and 36-month AFS probabilities.

Validation of nomogram performance

The total score of these predictors was 182.504 points. The optimal cut-off value of the nomogram was 89.026 in the primary cohort and 93.581 in the validation cohort. Under the optimal cut-off value, the SE, SP, PPV, NPV, PLR, and NLR were 69.70%, 57.41%, 50.00%, 75.61%, 1.636 and 0.528 respectively in the primary cohort, and were 31.03%, 86.96%, 60.00%, 66.67%, 2.379, and 0.793 respectively in the validation cohort.

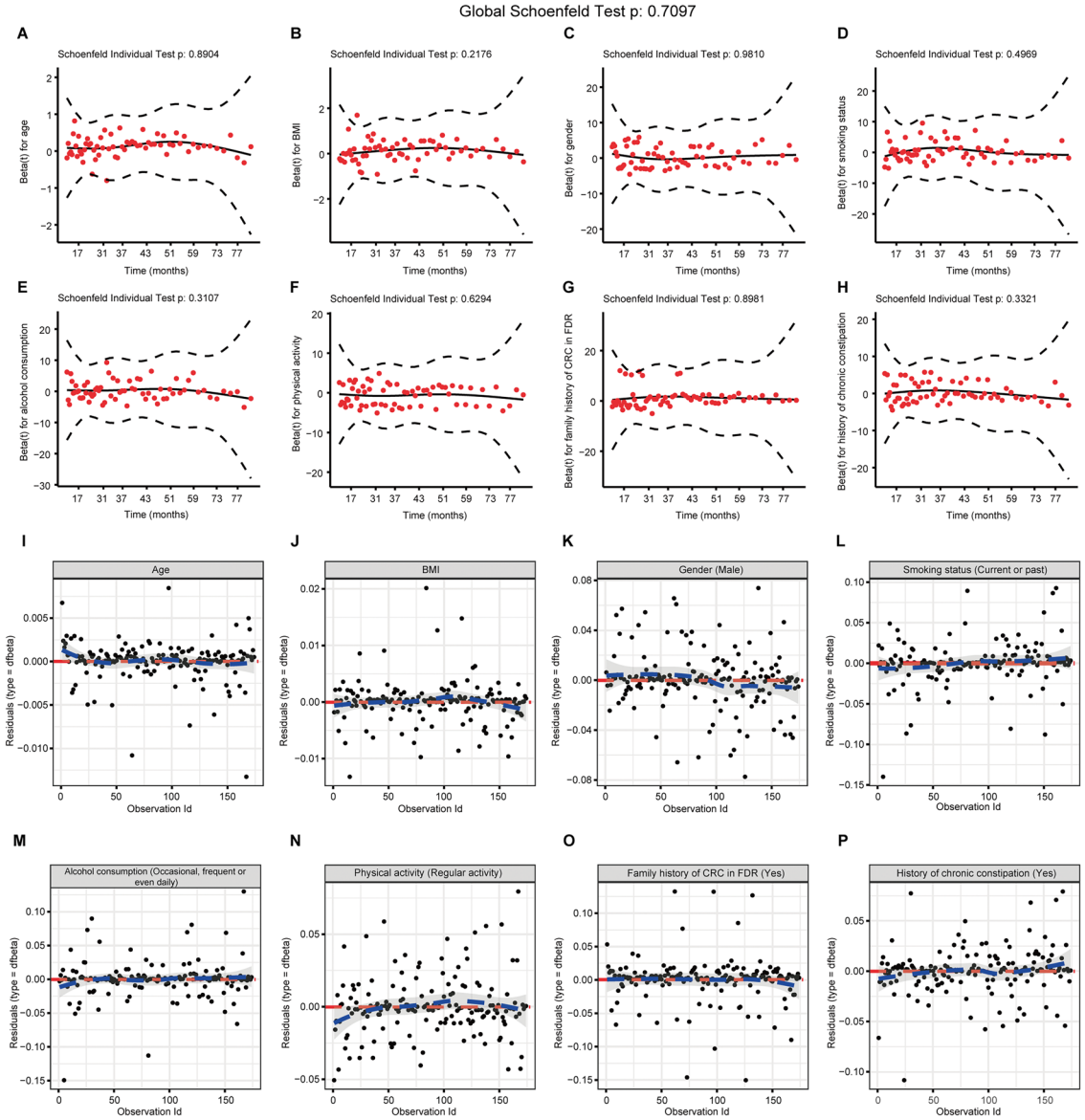


Figure 4. Schoenfeld's individual and global test and deviance residual were conducted to appraise whether the eight covariates selected using the LASSO regression were associated with time before performing the multivariate Cox-PH regression analysis. (**A**–**H**) Plots of the scaled Schoenfeld residuals against the changed time. The solid line represents a smoothing spline fit to the plot, and the dashed lines represent a ± 2-SE band around the fit. Considerable departures from the horizontal line indicate nonproportional hazards. (**I**–**P**) Index graphs of dfbeta for the Cox-PH regression of AFS. The graphs comparing the magnitudes of the largest dfbeta values with the regression coefficients uncover that there exist no greatly influential individual observations. AFS: adenoma-free survival.

To further evaluate the performance of Nomogram-1, the time-dependent ROC curves of the primary cohort and the validation cohort were delineated for AFS status (Fig. 6A). The C-index and bootstrapping-corrected C-index were calculated as 0.682 and 0.663 respectively for the primary cohort, indicating reliable predictive accuracy of the nomogram. The ROC curves of the primary cohort for 24-month, 30-month and 36-month AFS were shown in Fig. 6B–D. Calibration plots exhibited a robust pertinence between the actual probability (y-axis) and the predicted probability (x-axis) of 24-month AFS in the primary cohort and the validation cohort. There also existed a high consistency between the actual probability (y-axis) and the predicted probability (x-axis) of 30-month AFS in the primary cohort and the validation cohort (Fig. 6E–H). DCA for 24-month and 30-month AFS also manifested that applying our nomogram to identify patients who developed colorectal adenomas after negative index colonoscopy had an edge over the scheme of "surveillance colonoscopy for no patients" and the strategy of "surveillance colonoscopy for all patients", suggesting great clinical utility of the nomogram in both cohorts (Fig. 6I,J). To assess the risk stratification ability of Nomogram-1, participants were divided into the high-risk group and the low-risk group based on the optimal cut-off value of risk scores in each cohort. The KM survival curves with individual survival numbers and time data were delineated in

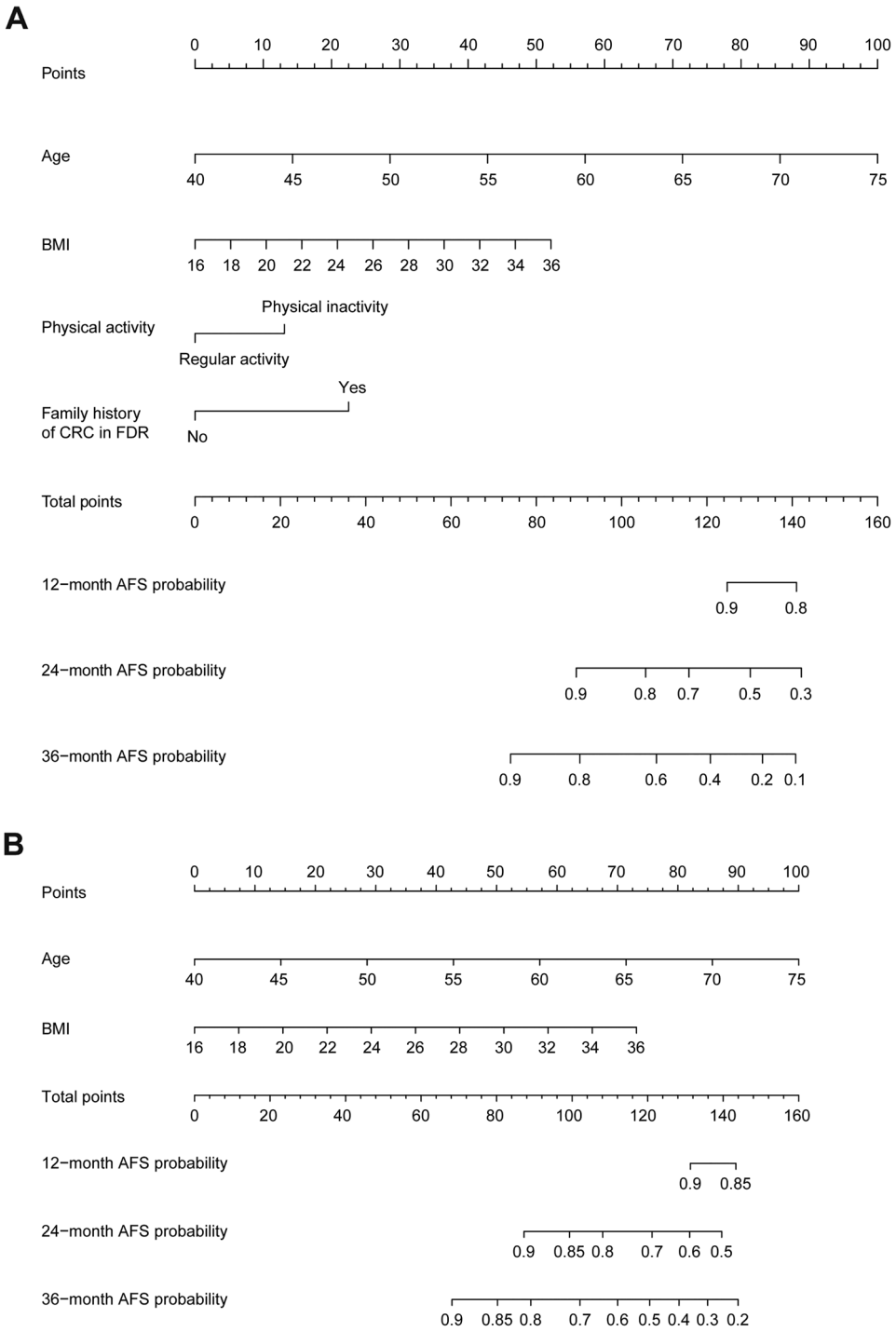


Figure 5. Construction of two nomogram models for predicting 12-month, 24-month, and 36-month AFS probabilities.

Fig. 6K–L. Log-rank tests revealed a significantly lower proportion of adenoma-free subjects in the high-risk group compared with that in the low-risk group in the primary cohort (Fig. 6K, $p < 0.0001$) and the validation cohort (Fig. 6L, $p = 0.017$).

Discussion

In this retrospective cohort study, we evaluated the pertinence between baseline patient characteristics involving modifiable lifestyle factors and adenoma occurrence risks based on high-risk populations with negative index colonoscopy. Four factors encompassing baseline age, BMI, physical activity and family history of CRC in FDR

Variables	Nomogram model-1			Nomogram model-2		
	HR[a] (95% CI)	*p*-Value	VIF	HR[b] (95% CI)	*p*-Value	VIF
Age	1.141 (1.078–1.207)	< 0.0001	1.106	1.108 (1.053–1.166)	< 0.0001	1.000
BMI	1.128 (1.029–1.237)	0.010	1.008	1.141 (1.040–1.251)	0.005	1.000
Physical activity		0.029			–	
Physical inactivity	Reference		Reference	–		–
Regular activity	0.547 (0.317–0.942)		1.062	–		–
Family history of CRC in FDR		0.016			–	
No	Reference		Reference	–		–
Yes	2.825 (1.214–6.572)		1.060	–		–

Table 2. Results of the multivariate Cox-PH regression analysis for construction of two nomogram models. [a]HRs were adjusted for baseline age, BMI, physical activity and family history of CRC in FDR. [b]HRs were adjusted for baseline age and BMI. *Cox-PH* Cox proportional hazards, *HR* hazard ratio, *CI* confidence interval, *VIF* variance inflation factor.

Variables	The univariate Cox-PH regression analysis		Variables	The multivariate Cox-PH regression analysis		
	HR (95% CI)	*p*-Value		HR[a] (95% CI)	*p*-Value	VIF
Age	1.099 (1.048–1.153)	0.001	Age	1.102 (1.047–1.161)	< 0.0001	1.040
BMI	1.147 (1.038–1.269)	0.007	BMI	1.133 (1.031–1.246)	0.010	1.003
Gender		0.025	Gender		0.272	
Female	Reference		Female	Reference		Reference
Male	1.743 (1.073–2.832)		Male	1.325 (0.802–2.189)		1.043
Smoking status		0.197	Smoking status		–	
Never	Reference		Never	–		–
Current or past	1.481 (0.816–2.691)		Current or past	–		–
Alcohol consumption		0.077	Alcohol consumption		–	
Never	Reference		Never	–		–
Occasional, frequent or even daily	1.683 (0.945–3.000)		Occasional, frequent or even daily	–		–
Physical activity		0.438	Physical activity		–	
Physical inactivity	Reference		Physical inactivity	–		–
Regular activity	0.822 (0.500–1.349)		Regular activity	–		–
Family history of CRC in FDR		0.255	Family history of CRC in FDR		–	
No	Reference		No	–		–
Yes	1.590 (0.716–3.529)		Yes	–		–
History of chronic constipation		0.495	History of chronic constipation		–	
No	Reference		No	–		–
Yes	1.203 (0.707–2.047)		Yes	–		–
History of chronic diarrhea		0.586	History of chronic diarrhea		–	
No	Reference		No	–		–
Yes	0.861 (0.502–1.477)		Yes	–		–
History of chronic appendicitis or appendectomy		0.704	History of chronic appendicitis or appendectomy		–	
No	Reference		No	–		–
Yes	1.179 (0.504–2.759)		Yes	–		–

Table 3. Variable selection through the univariate/multivariate Cox-PH regression analysis for predicting risks of colorectal adenoma occurrence based on the primary cohort. [a]HRs were adjusted for baseline age, BMI and gender.

were selected to construct a nomogram for predicting AFS probabilities. Our nomogram had great discriminatory ability for predicting AFS rates, with a C-index of 0.682 and a bootstrapping-corrected C-index of 0.663 in the primary cohort. Additionally, the calibration curves illustrated that the nomogram-predicted probability was closely aligned with the actual probability in both cohorts. The DCA curves and KM survival curves also illuminated that our model had sufficient net benefit for clinical application and convincing risk stratification ability.

Mounting studies have investigated putative risk factors for colorectal adenoma occurrence. Population-based national cancer screening programs emphasized that the detection rates for non-advanced adenomas, advanced

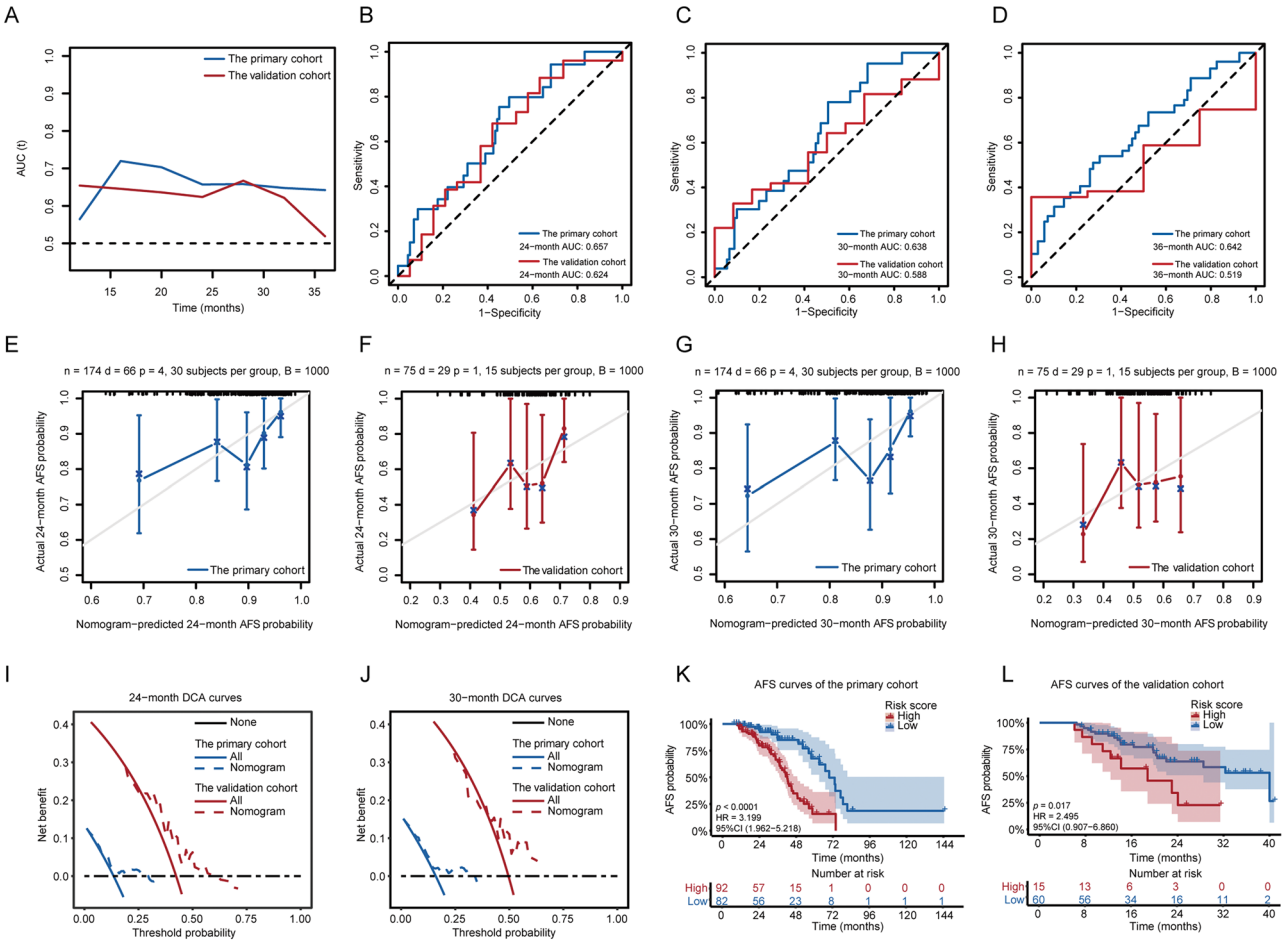


Figure 6. The predictive accuracy, discriminatory ability, clinical utility, and risk stratification capacity of the nomogram for predicting AFS probabilities were evaluated using the tROC curves, calibration plots, DCA curves, and KM survival curves. (**A**) The tROC curves indicated stable predictive accuracy of the nomogram over time in the primary cohort (blue) and the validation cohort (red). (**B**–**D**) The ROC curves of the nomogram for predicting 24-month (**B**), 30-month (**C**), and 36-month (**D**) AFS probabilities based on the primary cohort (blue) and the validation cohort (red). (**E**–**H**) Calibration plots exhibited a robust pertinence between the actual probability (y-axis) and the predicted probability (x-axis) of 24-month AFS in the primary cohort (**E**) and the validation cohort (**F**). There also existed a high consistency between the actual probability (y-axis) and the predicted probability (x-axis) of 30-month AFS in the primary cohort (**G**) and the validation cohort (**H**). The grey line represents the ideal fit. The blue or red line reflects the nomogram prediction, of which a closer fit to the grey line suggests better performance. (**I**,**J**) DCA of the nomogram in the primary cohort (blue) and the validation cohort (red) at 24-month (**I**) and 30-month (**J**) follow up. The black dotted line represents the screen-none scheme. The red or blue solid line represents the screen-all strategy. The red or blue dotted line represents the nomogram. (**K**,**L**) KM survival curves demonstrating the AFS probabilities in the primary cohort (**K**) and the validation cohort (**L**) with individual survival numbers and time data. tROC: time-dependent receiver operating characteristic; DCA: decision curve analysis; AUC: area under the curve.

neoplasms, and any neoplasms all increased with age[7]. A Korean study involving asymptomatic participants who underwent screening colonoscopies from January 2006 to June 2009 uncovered that the prevalence of advanced adenomas escalated from 0.60% among participants ages 30 to 39 to 7.40% among those ages 70–79[40]. Moreover, a 5-unit increase in BMI is correlated with a 19% increased risk of colorectal adenomas as demonstrated by a comprehensive meta-analysis[32]. Physical activity, especially leisure and sport activity, was corroborated to protect against colorectal adenomas[19]. A multinational, prospective colonoscopy study which involved 16 Asia–Pacific regions from December 2011 to December 2013 disclosed that subjects without a family history of CRC in FDR were less likely to have colorectal adenomas than those with at least one FDR affected (adjusted odds ratio = 1.31–1.92)[34]. In line with previous literature, older age and higher BMI were identified as independent risk factors for the development of colorectal adenomas by both the LASSO-Cox regression and the univariate/multivariate regression in the present study. Physical activity and family history of CRC in FDR were also independently associated with colorectal adenoma occurrence according to the LASSO-Cox regression analysis.

Moreover, the relationship between other studied factors and colorectal adenoma occurrence has also been systematically interrogated. Recent studies manifested that the prevalence and incidence of colorectal adenomas in women were significantly lower than those in men[30,33], possibly owing to the lower colonoscopy completion

rate among women[41], who prefer more convenient and non-invasive screening approaches such as computed tomography colonography[42] and multitarget stool DNA testing[43]. Additionally, previous studies disclosed that the prevalence of colorectal polyps increased by 3.40 times in current smokers as compared to never smokers[44]. Dose-response relations were also detected among the pack-years of smoking, the duration of smoking, the daily number of cigarettes smoked, and risks of colorectal adenomas[45]. DNA hypomethylation in the normal rectal mucosa[46], genetic variants in mismatch repair enzymes[47], and polymorphisms in tobacco-carcinogen-metabolizing pattern[48] might be parts of underlying mechanisms. Besides, recent meta-analysis underscored the positive correlation between alcohol consumption and risks of colorectal adenomas[18]. Proposed pathophysiological mechanisms involve the effects of acetaldehyde and reactive oxygen species, induction of cytochrome P 4502E1, as well as nutritional deficiency[49]. Furthermore, higher prevalence and incidence of CRC and benign colorectal neoplasms were observed among patients with chronic constipation compared with matched chronic constipation-free patients[35]. Chronic diarrhea was also considered as an independent risk factor for colorectal ademoma occurrence[31], presumably owing to structural and functional disruption of the colonic epithelium caused by increased levels of secondary bile acids through oxidative damage to DNA, inflammation, and enhanced cell proliferation[50]. A history of chronic appendicitis or appendectomy had a higher sensitivity than a history of adenomatous polyps for surveillance of colorectal adenomas[36]. However, no significant association between each of these variables and risks of colorectal ademoma occurrence was found in the present study, which might be attributed to different study designs and effects of small sample size.

Length of screening intervals and numbers of necessary examinations are crucial parameters for CRC screening. Considering the astronomical economic and healthcare burden of CRC screening and surveillance, greater efforts are warranted to ensure that colonoscopy services are delivered to qualified patients at proper intervals[10]. Hitherto, there is a lack of adequate and clear guidance on the timing of follow-up colonoscopies after negative index colonoscopy. For asymptomatic average-risk individuals whose index colonoscopy detected no polyps or detected merely hyperplastic polyps in the rectum and sigmoid colon sized < 10 mm, current guidelines generally recommend them to undergo surveillance colonoscopies after ten years[10,29,51]. As for high-risk adults with negative index colonoscopy, surveillance intervals recommended by current guidelines are broad and diverse, and most of them merely consider risk factors including family history of CRC in FDR, inflammatory bowel disease and hereditary syndromes[51–54] and tend to ignore modifiable lifestyle factors. Nevertheless, familial and hereditary factors and lifestyle-related patterns often co-exist and interact in the etiology of colorectal neoplasms[55], and the synergistic effects of multiple lifestyle factors are non-negligible in the scheduling of surveillance colonoscopies in real life[56,57]. This research gap may give rise to inaccurate risk predictions and inappropriate surveillance recommendations, thus raising the likelihood of over-use or under-use of surveillance colonoscopies. Over-use of colonoscopies is notorious for increased risks of adverse events, extra financial burdens, and deprived opportunities for under-screened populations to undergo adequate colonoscopy service, while under-use of colonoscopies is detrimental to the accuracy and timeliness of CRC diagnosis[10]. Notably, our model can provide accurate and convenient predictions of 24-month, 30-month and 36-month AFS probabilities for high-risk individuals with negative index colonoscopy based on their lifestyle patterns, thus optimizing the timing of follow-up colonoscopies for effective prevention of colorectal adenomas and CRC.

Specific limitations deserve careful attention during interpretation of our results. First, considering the retrospective design of our study and the subjective nature of lifestyle patterns, the inherent recall bias was inevitable. Second, conducted in a single center based on a relatively small sample of participants in Tianjin, our study was subjected to unavoidable selection bias, and the generalizability of our nomogram would be compromised correspondingly. Multicenter, large-scale, prospective investigations with external validation data from other cities and nations are warranted to polish our model in the future. Third, although we used the multivariate Cox-PH regression to control confounding factors, other factors such as dietary factors including intakes of red meat or processed meat, fiber and calcium, and some clinical features encompassing use of insulin, C-peptide and aspirin, were not controlled because related information was incomplete or missing. The most highlighted strength of our study is that the studied covariates are usually available from electronic medical records, thus ensuring the feasibility and convenience of applying our nomogram model to clinical practice. Furthermore, we excluded subjects whose first surveillance colonoscopy was conducted within six months after negative index colonoscopy or failed to meet any of the standards of qualified colonoscopies, so as to mitigate the impacts of missing adenomas on our analysis.

Conclusions

To summarize, we incorporated four baseline patient characteristics including age, BMI, physical activity and family history of CRC in FDR, which were selected using the LASSO-Cox regression analysis, into a nomogram model for predicting risks of colorectal adenoma occurrence among high-risk populations with negative index colonoscopy. Our model not only had good predictive accuracy, discriminatory ability and clinical application value, but also effectively stratified individuals with different risk scores into different risk tiers of survival outcomes. Our model will add a new dimension to the implementation of individualized prevention against colorectal adenomas and CRC.

Data availability

The datasets generated during and/or analysed during the current study are not publicly available due to protection of patients' privacy but are available from the corresponding author on reasonable request.

Received: 11 June 2023; Accepted: 15 May 2024
Published online: 21 May 2024

References

1. Sung, H. *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249. https://doi.org/10.3322/caac.21660 (2021).
2. Chan, A. T. & Giovannucci, E. L. Primary prevention of colorectal cancer. *Gastroenterology* **138**, 2029-2043.e2010. https://doi.org/10.1053/j.gastro.2010.01.057 (2010).
3. Bray, F. Transition in human development and the global cancer burden. (2014).
4. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–767. https://doi.org/10.1016/0092-8674(90)90186-i (1990).
5. Conteduca, V., Sansonno, D., Russi, S. & Dammacco, F. Precancerous colorectal lesions (Review). *Int. J. Oncol.* **43**, 973–984. https://doi.org/10.3892/ijo.2013.2041 (2013).
6. Siegel, R. L. *et al.* Colorectal cancer statistics, 2020. *CA Cancer J. Clin.* **70**, 145–164. https://doi.org/10.3322/caac.21601 (2020).
7. Chen, H. *et al.* Participation and yield of a population-based colorectal cancer screening programme in China. *Gut* **68**, 1450–1457. https://doi.org/10.1136/gutjnl-2018-317124 (2019).
8. Li, N. *et al.* Incidence, mortality, survival, risk factor and screening of colorectal cancer: A comparison among China, Europe, and northern America. *Cancer Lett.* **522**, 255–268. https://doi.org/10.1016/j.canlet.2021.09.034 (2021).
9. Bénard, F., Barkun, A. N., Martel, M. & von Renteln, D. Systematic review of colorectal cancer screening guidelines for average-risk adults: Summarizing the current global recommendations. *World J. Gastroenterol.* **24**, 124–138. https://doi.org/10.3748/wjg.v24.i1.124 (2018).
10. Murphy, C. C., Sandler, R. S., Grubber, J. M., Johnson, M. R. & Fisher, D. A. Underuse and overuse of colonoscopy for repeat screening and surveillance in the veterans health administration. *Clin. Gastroenterol. Hepatol.* **14**, 436-444.e431. https://doi.org/10.1016/j.cgh.2015.10.008 (2016).
11. Steinwachs, D. *et al.* National institutes of health state-of-the-science conference statement: Enhancing use and quality of colorectal cancer screening. *Ann. Intern. Med.* **152**, 663–667. https://doi.org/10.7326/0003-4819-152-10-201005180-00237 (2010).
12. Chen, H. *et al.* comparative evaluation of participation and diagnostic yield of colonoscopy vs fecal immunochemical test vs risk-adapted screening in colorectal cancer screening: Interim analysis of a multicenter randomized controlled trial (TARGET-C). *Am. J. Gastroenterol.* **115**, 1264–1274. https://doi.org/10.14309/ajg.0000000000000624 (2020).
13. Rex, D. K. *et al.* 5-year incidence of adenomas after negative colonoscopy in asymptomatic average-risk persons [see comment]. *Gastroenterology* **111**, 1178–1181. https://doi.org/10.1053/gast.1996.v111.pm8898630 (1996).
14. Neugut, A. I. *et al.* Incidence and recurrence rates of colorectal adenomas: A prospective study. *Gastroenterology* **108**, 402–408. https://doi.org/10.1016/0016-5085(95)90066-7 (1995).
15. Botteri, E., Iodice, S., Raimondi, S., Maisonneuve, P. & Lowenfels, A. B. Cigarette smoking and adenomatous polyps: A meta-analysis. *Gastroenterology* **134**, 388–395. https://doi.org/10.1053/j.gastro.2007.11.007 (2008).
16. Pan, J. *et al.* Prevalence and risk factors for colorectal polyps in a Chinese population: A retrospective study. *Sci. Rep.* **10**, 6974. https://doi.org/10.1038/s41598-020-63827-6 (2020).
17. Sandler, R. S., Lyles, C. M., McAuliffe, C., Woosley, J. T. & Kupper, L. L. Cigarette smoking, alcohol, and the risk of colorectal adenomas. *Gastroenterology* **104**, 1445–1451. https://doi.org/10.1016/0016-5085(93)90354-f (1993).
18. Zhu, J. Z. *et al.* Systematic review with meta-analysis: alcohol consumption and the risk of colorectal adenoma. *Aliment Pharmacol. Ther.* **40**, 325–337. https://doi.org/10.1111/apt.12841 (2014).
19. Sandler, R. S., Pritchard, M. L. & Bangdiwala, S. I. Physical activity and the risk of colorectal adenomas. *Epidemiology* **6**, 602–606. https://doi.org/10.1097/00001648-199511000-00007 (1995).
20. Wolin, K. Y., Yan, Y. & Colditz, G. A. Physical activity and risk of colon adenoma: A meta-analysis. *Br. J. Cancer* **104**, 882–885. https://doi.org/10.1038/sj.bjc.6606045 (2011).
21. Wu, W. M. *et al.* Colorectal cancer screening modalities in chinese population: Practice and lessons in Pudong new area of Shanghai, China. *Front. Oncol.* **9**, 399. https://doi.org/10.3389/fonc.2019.00399 (2019).
22. Balachandran, V. P., Gonen, M., Smith, J. J. & DeMatteo, R. P. Nomograms in oncology: More than meets the eye. *Lancet Oncol.* **16**, e173-180. https://doi.org/10.1016/s1470-2045(14)71116-7 (2015).
23. Saito, Y. *et al.* Colonoscopy screening and surveillance guidelines. *Dig. Endosc.* **33**, 486–519. https://doi.org/10.1111/den.13972 (2021).
24. Aronchick, C. A., Lipshutz, W. H., Wright, S. H., Dufrayne, F. & Bergman, G. A novel tableted purgative for colonoscopic preparation: Efficacy and safety comparisons with Colyte and Fleet Phospho-Soda. *Gastrointest Endosc.* **52**, 346–352. https://doi.org/10.1067/mge.2000.108480 (2000).
25. Barclay, R. L., Vicari, J. J., Doughty, A. S., Johanson, J. F. & Greenlaw, R. L. Colonoscopic withdrawal times and adenoma detection during screening colonoscopy. *N. Engl. J. Med.* **355**, 2533–2541. https://doi.org/10.1056/NEJMoa055498 (2006).
26. Harris, J. K. *et al.* Factors associated with the technical performance of colonoscopy: An EPAGE Study. *Dig. Liver Dis.* **39**, 678–689. https://doi.org/10.1016/j.dld.2007.02.012 (2007).
27. Huang, Y. *et al.* Recurrence and surveillance of colorectal adenoma after polypectomy in a southern Chinese population. *J. Gastroenterol.* **45**, 838–845. https://doi.org/10.1007/s00535-010-0227-3 (2010).
28. He, Q. *et al.* Development and validation of a nomogram based on neutrophil-to-lymphocyte ratio and fibrinogen-to-lymphocyte ratio for predicting recurrence of colorectal adenoma. *J. Gastrointest Oncol.* **13**, 2269–2281. https://doi.org/10.21037/jgo-22-410 (2022).
29. Shussman, N. & Wexner, S. D. Colorectal polyps and polyposis syndromes. *Gastroenterol. Rep. (Oxf.)* **2**, 1–15. https://doi.org/10.1093/gastro/got041 (2014).
30. Brenner, H., Altenhofen, L., Stock, C. & Hoffmeister, M. Incidence of colorectal adenomas: Birth cohort analysis among 4.3 million participants of screening colonoscopy. *Cancer Epidemiol. Biomarkers Prev.* **23**, 1920–1927. https://doi.org/10.1158/1055-9965.Epi-14-0367 (2014).
31. Li, W. *et al.* Establish a novel model for predicting the risk of colorectal Ademomatous polyps: A prospective cohort study. *J. Cancer* **13**, 3103–3112. https://doi.org/10.7150/jca.74772 (2022).
32. Ben, Q. *et al.* Body mass index increases risk for colorectal adenomas based on meta-analysis. *Gastroenterology* **142**, 762–772. https://doi.org/10.1053/j.gastro.2011.12.050 (2012).
33. McCashland, T. M., Brand, R., Lyden, E. & de Garmo, P. Gender differences in colorectal polyps and tumors. *Am. J. Gastroenterol.* **96**, 882–886. https://doi.org/10.1111/j.1572-0241.2001.3638_a.x (2001).
34. Wong, M. C. *et al.* Risk of colorectal neoplasia in individuals with self-reported family history: A prospective colonoscopy study from 16 asia-pacific regions. *Am. J. Gastroenterol.* **111**, 1621–1629. https://doi.org/10.1038/ajg.2016.52 (2016).
35. Guérin, A. *et al.* Risk of developing colorectal cancer and benign colorectal neoplasm in patients with chronic constipation. *Aliment Pharmacol. Ther.* **40**, 83–92. https://doi.org/10.1111/apt.12789 (2014).
36. Meng, W. *et al.* Performance value of high risk factors in colorectal cancer screening in China. *World J. Gastroenterol.* **15**, 6111–6116. https://doi.org/10.3748/wjg.15.6111 (2009).
37. Akaike, H. T. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **19**, 716–723 (1974).
38. Harrell, F. E. Jr., Califf, R. M., Pryor, D. B., Lee, K. L. & Rosati, R. A. Evaluating the yield of medical tests. *Jama* **247**, 2543–2546 (1982).

39. Marquardt, D. W. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics* **12**, 591–612. https://doi.org/10.2307/1267205 (1970).
40. Yang, M. H. *et al.* The prevalence of colorectal adenomas in asymptomatic Korean men and women. *Cancer Epidemiol. Biomarkers Prev.* **23**, 499–507. https://doi.org/10.1158/1055-9965.Epi-13-0682 (2014).
41. Etzioni, D. A. *et al.* A population-based study of colorectal cancer test use: Results from the 2001 california health interview survey. *Cancer* **101**, 2523–2532. https://doi.org/10.1002/cncr.20692 (2004).
42. Chu, L. L., Weinstein, S. & Yee, J. Colorectal cancer screening in women: An underutilized lifesaver. *AJR Am. J. Roentgenol.* **196**, 303–310. https://doi.org/10.2214/ajr.10.5815 (2011).
43. Imperiale, T. F. *et al.* Multitarget stool DNA testing for colorectal-cancer screening. *N. Engl. J. Med.* **370**, 1287–1297. https://doi.org/10.1056/NEJMoa1311194 (2014).
44. Lee, K. & Kim, Y. H. Colorectal polyp prevalence according to alcohol consumption, smoking and obesity. *Int. J. Environ. Res. Public Health.* **17**, 2387. https://doi.org/10.3390/ijerph17072387 (2020).
45. Shrubsole, M. J. *et al.* Alcohol drinking, cigarette smoking, and risk of colorectal adenomatous and hyperplastic polyps. *Am. J. Epidemiol.* **167**, 1050–1058. https://doi.org/10.1093/aje/kwm400 (2008).
46. Paun, B. C. *et al.* Relation between normal rectal methylation, smoking status, and the presence or absence of colorectal adenomas. *Cancer* **116**, 4495–4501. https://doi.org/10.1002/cncr.25348 (2010).
47. Yu, J. H., Bigler, J., Whitton, J., Potter, J. D. & Ulrich, C. M. Mismatch repair polymorphisms and colorectal polyps: hMLH1–93G>A variant modifies risk associated with smoking. *Am. J. Gastroenterol.* **101**, 1313–1319. https://doi.org/10.1111/j.1572-0241.2006.00551.x (2006).
48. Fu, Z. *et al.* Interaction of cigarette smoking and carcinogen-metabolizing polymorphisms in the risk of colorectal polyps. *Carcinogenesis* **34**, 779–786. https://doi.org/10.1093/carcin/bgs410 (2013).
49. Seitz, H. K., Maurer, B. & Stickel, F. Alcohol consumption and cancer of the gastrointestinal tract. *Dig. Dis.* **23**, 297–303. https://doi.org/10.1159/000090177 (2005).
50. Jia, W., Xie, G. & Jia, W. Bile acid-microbiota crosstalk in gastrointestinal inflammation and carcinogenesis. *Nat. Rev. Gastroenterol. Hepatol.* **15**, 111–128. https://doi.org/10.1038/nrgastro.2017.119 (2018).
51. Provenzale, D. *et al.* NCCN guidelines insights: Colorectal cancer screening, version 1.2018. *J. Natl. Compr. Canc. Netw.* **16**, 939–949. https://doi.org/10.6004/jnccn.2018.0067 (2018).
52. Choi, Y., Sateia, H. F., Peairs, K. S. & Stewart, R. W. Screening for colorectal cancer. *Semin. Oncol.* **44**, 34–44. https://doi.org/10.1053/j.seminoncol.2017.02.002 (2017).
53. Syngal, S. *et al.* ACG clinical guideline: Genetic testing and management of hereditary gastrointestinal cancer syndromes. *Am. J. Gastroenterol.* **110**, 223–262. https://doi.org/10.1038/ajg.2014.435 (2015).
54. Weiss, J. M. *et al.* NCCN guidelines® insights: Genetic/familial high-risk assessment: Colorectal, version 1.2021. *J. Natl. Compr. Canc. Netw.* **19**, 1122–1132. https://doi.org/10.1164/jnccn.2021.0048 (2021).
55. Sawicki, T. *et al.* A review of colorectal cancer in terms of epidemiology, risk factors, development, symptoms and diagnosis. *Cancers (Basel).* **13**, 2025. https://doi.org/10.3390/cancers13092025 (2021).
56. Fu, Z. *et al.* Lifestyle factors and their combined impact on the risk of colorectal polyps. *Am. J. Epidemiol.* **176**, 766–776. https://doi.org/10.1093/aje/kws157 (2012).
57. Tabung, F. K. *et al.* A healthy lifestyle index is associated with reduced risk of colorectal adenomatous polyps among non-users of non-steroidal anti-inflammatory drugs. *J. Prim. Prev.* **36**, 21–31. https://doi.org/10.1007/s10935-014-0372-1 (2015).

Acknowledgements

This research was supported by the Key R&D Projects in the Tianjin Science and Technology Pillar Program (Grant number 19YFZCSY00420), Natural Science Foundation of Tianjin (21JCZDJC00060, 21JCYBJC00180 and 21JCYBJC00340), Tianjin Key Medical Discipline (Specialty) Construction Project (Grant number TJYXZDXK-044A) and Tianjin Hospital Association Hospital Management Research Project (Grant number 2019ZZ07).

Author contributions

C.Z., X.Z. and Q.Z. conceived and designed the study. M.Y., Y.OY., W.P., S.Y., X.L., W.W. and B.Y. perused electronic medical records and collected updated data concerning studied risk factors. M.Y., Y.OY., Z.Y. and S.W. took charge of data disposal, result interpretation and drafting the main body of the manuscript. Q.H., Y.Y., Y.L. were involved in depicting the figures. J.S., T.C. and Z.F. participated in drawing the tables. C.Z. and X.Z. made critical revision of the manuscript for significant intellectual content. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-62348-w.

Correspondence and requests for materials should be addressed to X.Z. or C.Z.