Cancer Management and Research

Dovepress

open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Assessing Individual Risk for High-Risk Early Colorectal Neoplasm for Pre-Selection of Screening in Shanghai, China: A Population-Based Nested Case–Control Study

Jie Shen[1,2,*]
Yiling Wu[3,*]
Xiaoshuang Feng[1,2,*]
Fei Liang[4]
Miao Mo[1,2]
Binxin Cai[3]
Changming Zhou[1,2]
Zezhou Wang[1,2]
Meiying Zhu[3]
Guoxiang Cai[2,5]
Ying Zheng[1,2]

[1]Department of Cancer Prevention, Fudan University Shanghai Cancer Center, Shanghai, People's Republic of China; [2]Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, People's Republic of China; [3]Department of Noninfectious Chronic Disease Control and Prevention, Songjiang District Center for Disease Control and Prevention, Shanghai, People's Republic of China; [4]Department of Biostatistics, Zhongshan Hospital Fudan University, Shanghai, People's Republic of China; [5]Department of Colorectal Surgery, Fudan University Shanghai Cancer Center, Shanghai, People's Republic of China

*These authors contributed equally to this work

Correspondence: Ying Zheng
Department of Cancer Prevention, Fudan University Shanghai Cancer Center, No. 270 Dong'An Road, Shanghai, 200032, People's Republic of China
Email zhengying@fudan.edu.cn

Objective: To identify people with high-risk early colorectal neoplasm is highly desirable for pre-selection in colorectal cancer (CRC) screening in low-resource countries. We aim to build and validate a risk-based model so as to improve compliance and increase the benefits of screening.

Patients and Methods: Using data from the Shanghai CRC screening cohort, we conducted a population-based nested case–control study to build a risk-based model. Cases of early colorectal neoplasm were extracted as colorectal adenomas and stage 0-I CRC. Each case was matched with five individuals without neoplasm (controls) by the screening site and year of enrollment. Cases and controls were then randomly divided into two groups, with two thirds for building the risk prediction model and the other one third for model validation. Known risk factors were included for risk prediction models using logistic regressions. The area under the receiver operating characteristic curve (AUC) and Hosmer–Lemeshow chi-square statistics were used to evaluate model discrimination and calibration. The predicted individual risk probability was calculated under the risk regression equation.

Results: The model incorporating age, sex, family history and lifestyle factors including body mass index (BMI), smoking status, alcohol, regular moderate-to-intensity physical activity showed good calibration and discrimination. When the risk cutoff threshold was defined as 17%, the sensitivity and specificity of the model were 63.99% and 53.82%, respectively. The validation data analysis also showed well discrimination.

Conclusion: A risk prediction model combining personal and lifestyle factors was developed and validated for high-risk early colorectal neoplasm among the Chinese population. This risk-based model could improve the pre-selection for screening and contribute a lot to efficient population-based screening in low-resource countries.

Keywords: cancer prevention, colorectal cancer, risk model, screening

Introduction

Colorectal cancer (CRC) is the fourth most common cancer and the second leading cause of cancer death globally.[1] A total of 1,849,518 new cancer cases and 880,792 cancer deaths occurred in 2018 worldwide, according to the estimation of GLOBOCAN series of the International Agency for Research on Cancer (IARC).[1,2] In China, CRC was the second leading cause of cancer death, and its incidence has increased rapidly over the last two decades.[3,4]

Received: 13 January 2021
Accepted: 2 April 2021
Published: 12 May 2021

Substantial evidences have demonstrated that CRC screening could help detect CRC in an earlier stage or a premalignant phase so as to reduce mortality and incidence substantially.[5,6] The incidence and mortality rates of CRC have been declining for several decades in the USA because of the introduction of screening and improvements in treatment.[7] Many countries and areas in the world have thus initiated CRC screening programs.

Shanghai launched a large-scale CRC screening program from 2012 for residents aged 50–74 years, which achieved significant progress, especially on the early detection of CRC and colorectal adenomas.[8] However, the target population of the current screening guideline comprises average-risk individuals, resulting in a low participation rate of screening. The compliance to colonoscopy in the Shanghai CRC screening program was 39.8%,[8] much lower than that in the USA[9] (53%) and Norway[10] (60.7%). Data showed that as high as 60%-70% of participants in urban China who were positive in initial testing refused further colonoscopy detection for various reasons.[11] The acceptance of colonoscopy in the general population is not optimal and is still a major challenge.[12] Other factors such as lack of knowledge about CRC and screening, psychological factors such as fear of screening complications or discomfort, distrust of community hospitals, insufficient advocacy by healthcare professionals and high cost of screening are also barriers to colonoscopy in China.[13,14]

Moreover, only around 10% of the average-risk people who underwent colonoscopy were diagnosed with adenoma or cancer.[11,15] Suboptimal screening of people at high risk greatly limited the early detection of colorectal neoplasm, delayed the diagnosis and treatment and hindered the screening. Therefore, the compliance of screening needs to be improved, and more screening benefits should be brought to this population, especially in low-resource countries. Besides, based on the adenoma-carcinoma sequence, the risk prediction for early colorectal neoplasm defined as colorectal adenomas and early-stage CRC may be more informative for screening and conservative for cancer prevention.

The risk stratification of target populations to be screened may bring more benefit and make screening more cost-effective. Numerous models have been developed to determine the risk of CRC,[16–28] and advanced colorectal adenomas (ACN),[29–34] most of which included clinical, lifestyle, and environmental risk factors; some even included genetic factors. The majority of the models were developed among Asians, US adults and Europeans in high-resource countries. Only three models for Chinese individuals,[20,27,31] however, were built from hospital-based medical records[20,31] or considering the history of clinical symptoms,[27] resulting in selection bias. No prediction tools specific to the Chinese population have been developed based on the large population-based data in mainland China.

This population-based nested case–control study was performed to develop a risk prediction model based on the data from the first round of Shanghai CRC screening. Using this risk stratification tool, the study aimed to identify individuals at an increased risk of early colorectal neoplasm, so as to provide evidence for effective screening strategy.

Patients and Methods

Study Population

The data were obtained from residents of the local Songjiang District community in Shanghai, east of China. Songjiang District is located in the north-west suburban area of Shanghai, with an area of 464.2 km^2 and a population of 1,568,000 residents in 2015; it represents characteristics of the typical urban-rural-integration-area of the Chinese population. The Songjiang local CRC screening cohort was initially established in 2012 and covered all native residents by the end of 2016.

The detailed screening protocol has been described previously.[8,15] In brief, residents aged 50–74 years with no CRC history were invited for participation. The screening consisted of two steps: initial fecal immunochemical tests (FIT) and risk assessment, followed by diagnostic colonoscopy testing for those with positive results. Participants who met the following criteria were included in this cohort: (1) residents aged 50–74 years; (2) residents who signed informed consent; (3) residents completing the questionnaire; and (4) residents having no personal history of CRC. A total of 176,432 residents participated in the first round of screening from 2012 to 2016. Further, 8534 participants without baseline risk information, 2035 participants with a CRC history, and 23,685 participants younger than 50 years or older than 74 years were excluded. Finally, the study cohort comprised 142,178 individuals.

A nested case–control study was then conducted using the screening cohort data. All the cases of early colorectal neoplasm defined as colorectal adenomas and stage

0-I CRC were included as cases. For each case, five controls were chosen from cohort participants alive and free of colorectal neoplasm, individually matched to cases by the screening site and year of enrollment. Thus, 11,076 individuals (1846 cases and 9230 matched controls) were included in the study. All the cases and matched controls were then randomly divided into two groups, two thirds for building risk prediction model (7752 individuals, including 1292 cases and 6460 controls) and the other one third for model evaluation (3324 individuals, including 554 cases and 2770 controls) (Figure 1).

All participants provided written informed consent. All participants were treated in accordance with the Declaration of Helsinki. Research ethics was waived by the institutional research committee of Shanghai Municipal Center of Disease Control and Prevention because it was a public program covering all citizens at a certain range of ages in Shanghai.

Outcome Ascertainment

The primary outcome was early colorectal neoplasm, including colorectal adenomas and stage 0-I CRC. Participants from 2012 were followed until the occurrence of colorectal neoplasm, death or December 31, 2016. CRC cases were defined as colorectal adenocarcinoma and confirmed by hospital medical records or pathology reports of colonoscopy. Colorectal adenomas were defined as histologically classified tubular adenoma, mixed adenoma, and villous adenoma and confirmed by histopathological reports of colonoscopy. Participants with cancer were tracked through CRC screening results and annual searches of the Shanghai cancer registry system. The outcome information was further confirmed by a medical record review performed by clinical experts.

Assessment of Risk Factors

Experienced primary care physicians conducted face-to-face interview using a standardized questionnaire, including information on demography and lifestyle factors.

Body weight and height were measured at baseline while subjects wore lightweight non-footwear. Body mass index (BMI) was calculated as follows: BMI = weight (kg)/height (m^2). BMI was categorized as normal weight: 18.5 ≤ BMI < 24 kg/m^2; slim: BMI <18.5 kg/m^2; overweight: 24 ≤ BMI < 28 kg/m^2; obesity: BMI ≥ 28 kg/m^2. The smoking status was classified as smokers (current/ever) or non-smokers. The alcohol drinking status was classified as drinkers (current/ever) or non-drinkers. The dietary factors, including the intake of fatty meat, fried food, and pickled food were categorized as occasional (<3 days per week) or regular (3–7 days per week) and moderate-to-intensity physical activity as never or regular (at least one day per week).

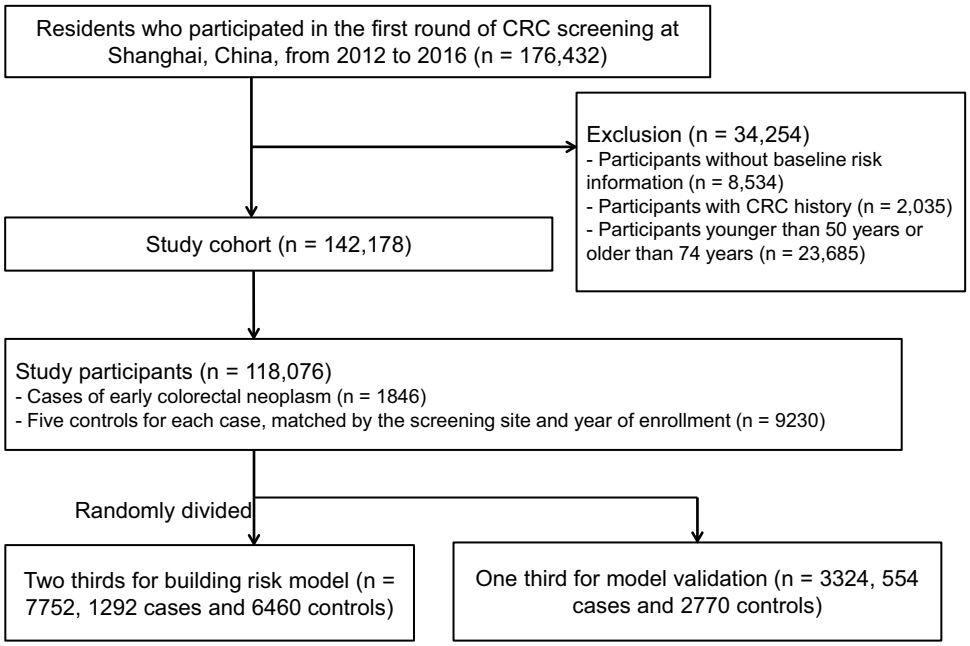

Figure 1 Flow diagram for selection of study cohort and nested case–control participants.

Statistical Analysis

The general characteristics of the participants were expressed as mean and standard deviation (SD) for continuous variables, and percentages for categorical variables. The data management and all analyses were performed using the SPSS 22.0 statistical software package (IBM, NY, USA).

Univariate analysis was carried out on the training dataset to examine the association between each risk factor and early colorectal neoplasm. Variables associated with early colorectal neoplasm in univariate analyses ($P < 0.05$) were entered in multivariate logistic regression models. Risk factors (variables) that retained significance in multivariate analyses were selected for incorporation into the risk model. Odds ratios (ORs) and 95% confidence intervals (CIs) of the association and the personal risk for early colorectal neoplasm were calculated. A risk equation for early colorectal neoplasm was created based on multivariate logistic regression with identified risk factors.

The calibration of the model was evaluated using the Hosmer-Lemeshow goodness-of-fit test, where a P value more than 0.05 indicated adequate calibration. The discrimination of the model was assessed using receiver operating characteristic (ROC) analysis. The area under the ROC curve (AUC) and 95% CIs were calculated, where the AUC value near 1 demonstrated excellent predictive ability.

The risk model was further evaluated in the independent validation dataset. With the same risk categories and equations in the training dataset, the personal risk for early colorectal neoplasm in the validation dataset was predicted and then the model was assessed by estimating the AUC value and 95% CI.

Results

Characteristics of the Training and Validation Sets

The characteristics of participants in the training and validation sets are listed in Table 1. Among 7752 participants in

Table 1 Characteristics of Participants in the Training and Validation Sets

	Training Set (n = 7752)	Validation Set (n = 3324)	P value
Age (year),mean (SD)	60.28 (6.34)	60.15 (6.41)	0.29
Sex, n (%)			0.44
Female	3931 (50.71)	1712 (51.50)	
Male	3821 (49.29)	1612 (48.50)	
BMI (kg/m^2), n (%)	23.88 (3.03)	23.90 (2.99)	0.62
18.5 ≤ BMI < 24	2876 (50.99)	1191 (50.17)	
BMI < 18.5	146 (2.59)	60 (2.53)	
24 ≤ BMI < 28	2128 (37.73)	930 (39.17)	
BMI ≥ 28	490 (8.69)	193 (8.13)	
Family history of CRC in first-degree relatives, n (%)	258 (3.34)	114 (3.44)	0.78
Smoking[a], n (%)	2636 (34.00)	1089 (32.81)	0.23
Alcohol[b], n (%)	1694 (21.86)	721 (21.75)	0.90
Regular intake of fatty food[c], n (%)	1670 (22.05)	725 (22.42)	0.67
Regular intake of fried food[d], n (%)	522 (7.04)	231 (7.28)	0.68
Regular intake of pickled food[e], n (%)	1835 (24.30)	762 (23.70)	0.52
Regular moderate-to-intensity physical activity[f], n (%)	5678 (81.84)	1936 (81.07)	0.41
Early colorectal neoplasm, n (%)			0.05
Adenomas	1122 (14.47)	493 (14.83)	
Stage 0–I colorectal cancer	170 (2.19)	61 (1.84)	

Notes: [a]Smoking status was classified as smokers (current/ever) or non-smoker. [b]Alcohol drinking status was classified as drinkers (current/ever) or non-drinker. [c]Regular intake of fatty meat means the intake 3–7 days per week [d]Regular intake of fried food means the intake 3–7 days per week. [e]Regular intake of pickled food means the intake 3–7 days per week. [f]Regular moderate-to-intensity physical activity means at least 1 day per week.

Abbreviations: BMI, body mass index; SD, standard deviation; CRC, colorectal cancer.

the training set, 1122 (14.47%) developed colorectal adenomas, and 170 (2.19%) were diagnosed with early-stage (stage 0-I) CRC. The mean age was 60.28 years (SD 6.34 years), 3821 (49.29%) were men, the obesity (BMI ≥ 28 kg/m^2) rate was 8.69%, and only 3.34% had a positive family history of first-degree relatives with CRC. Also, 34.00% of the participants were current/ever smokers and 21.86% were alcohol drinkers. The proportion of participants taking fatty food, fried food and pickled food regularly was 22.05%, 7.04%, and 24.30%, respectively, the rate of regular moderate-to-intensity physical activity every week was 81.84%.

The validation dataset was similar to the training set, with the mean age of 60.15 years (SD 6.41 years) and 114 participants (3.44%) had a family history of first-degree relatives with CRC among 3324 participants. Moreover, 493 (14.83%) colorectal adenomas and 61 (1.84%) 0-I stage CRC cases were diagnosed in the validation set.

Those who developed early colorectal neoplasms were significantly older, and more likely to have a positive family history of CRC compared with controls. In addition, cases were less likely to live a healthy lifestyle, with more smokers, more alcohol drinkers, and increased intake of fatty meat, fried food, and pickled food but less physical activities compared with controls (Table 2).

Univariate and Multivariate Predictors of Early Colorectal Neoplasm in the Training Dataset

Univariate and multivariate analyses were performed for the risk factors of early colorectal neoplasm. Multivariate logistic regression showed that age (60–69 years: OR = 1.45, 95% CI 1.25–1.69; ≥70 years: OR = 1.37, 95% CI 1.07–1.75), sex (male: OR = 1.35, 95% CI 1.09–1.67), a positive family history in first-degree relatives (OR = 2.78, 95% CI 2.08–3.71), obesity (OR = 1.51, 95% CI 1.18–1.92), smoking (OR = 1.28, 95% CI 1.04–1.58), and alcohol (OR = 1.23, 95% CI 1.02–1.48) were significant risk factors for early colorectal neoplasm, while regular moderate-to-intensity physical activity had an inverse association with the risk of early colorectal neoplasm (OR = 0.79, 95% CI 0.66–0.94). The adjusted predictors of early colorectal neoplasm are displayed in Table 3.

Development of the Risk Model and Assessment

The risk prediction model for early colorectal neoplasm included independent predictors as follows: age (categorized as 50–59 years, 60–69 years, and ≥70 years), sex (male/female), family history of CRC in first-degree relatives (yes/no), BMI (normal weight, slim, overweight, and obesity), smoker (smoker and non-smoker), alcohol (drinker and non-drinker), and moderate-to-intensity physical activity (never/regular).

The risk model equation was as follows:

$$\begin{aligned}\text{logit(P)} = &-1.985 + 0.374(\text{aged } 60-69)\\ &+ 0.312(\text{aged} \geq 70) + 0.300(\text{male})\\ &+ 0.097(\text{BMI}<18.5) + 0.119(24 \leq \text{BMI}<28)\\ &+ 0.411(\text{BMI} \geq 28)\\ &+ 1.021(\text{positve family history})\\ &+ 0.246(\text{smoker}) + 0.203(\text{alcohol})\\ &- 0.235(\text{regular exercise})\end{aligned}$$

The predicted risk probability equation was as follows:

$$\text{P} = \frac{EXP(Logit(P))}{1 + EXP(Logit(P))}$$

The risk probability of each individual participant was calculated based on the prediction model.

The Hosmer-Lemeshow goodness-of-fit test was used to evaluate the reliability of the model in the training dataset and P value of 0.94 indicated a good match of predicted risk over observed risk. Moreover, the model showed good discrimination with an AUC estimate of 0.624 (95% CI, 0.604–0.643). The model including lifestyle factors was significantly better than the model with only age and family history (Figure 2).

The risk prediction model was further assessed in the validation group. The risks in 3324 participants were predicted and the AUC estimate was 0.630 (95% CI, 0.604–0.655), showing high consistency with the training dataset.

The discriminative performances at the specific cutoff threshold of each predicted risk probability are provided in Table 4. With the highest Youden's index of 0.17, a risk cutoff threshold of 17% yielded a sensitivity of 63.99% and a specificity of 53.82%. At cutoffs yielding 90% specificity, the sensitivity for detecting early colorectal neoplasm was 21.2%.

Table 2 Baseline Characteristics of Cases of Early Colorectal Neoplasm and Controls in the Training and Validation Datasets

	Training Data		P value	Validation Data		P value
	Cases (n = 1292)	Controls (n = 6460)		Cases (n = 554)	Controls (n = 2770)	
Age,year,mean (SD)	60.90 (5.99)	60.16 (6.40)	0.00	61.1 (5.99)	59.95 (6.47)	0.00
Sex, n(%)			0.00			0.00
Male	799 (61.84)	3022 (46.78)		345 (62.27)	1267 (45.74)	
Female	493 (38.16)	3438 (53.20)		209 (37.73)	1503 (54.26)	
BMI, kg/m,^{2} mean (SD)	24.19 (3.09)	23.81 (3.02)	0.00	24.20 (3.19)	23.84 (2.95)	0.02
18.5 ≤ BMI < 24	485 (47.41)	2391 (51.79)		204 (48.11)	987 (50.62)	
BMI < 18.5	24 (2.35)	122 (2.64)		6 (1.42)	54 (2.77)	
24 ≤ BMI < 28	397 (38.81)	1731 (37.49)		175 (41.27)	755 (38.72)	
BMI ≥ 28	117 (11.44)	373 (8.08)		39 (9.20)	154 (7.90)	
Family history, n (%)			0.00			0.00
Yes	89 (6.92)	169 (2.62)		51 (9.24)	63 (2.28)	
No	1197 (93.08)	6270 (97.38)		501 (90.76)	2698 (97.72)	
Smoking, n (%)			0.00			0.00
No	708 (54.80)	4408 (68.24)		310 (55.96)	1920 (69.44)	
Current/ever	574 (45.20)	2052 (31.76)		244 (44.04)	845 (30.56)	
Alcohol, n (%)			0.00			0.00
No	893 (69.12)	5161 (79.94)		384 (69.44)	2210 (80.01)	
Current/ever	399 (30.88)	1295 (20.06)		169 (30.56)	552 (19.99)	
Intake of fatty food[a], n (%)			0.05			0.96
Occasional	964 (75.85)	4941 (78.38)		417 (77.51)	2091 (77.59)	
Regular	307 (24.15)	1363 (21.62)		121 (22.49)	604 (22.41)	
Intake of fried food[a], n (%)			0.02			0.31
Occasional	1135 (91.38)	5754 (93.27)		486 (93.82)	2455 (92.50)	
Regular	107 (8.62)	415 (6.73)		32 (6.18)	199 (7.50)	
Intake of pickled food[a], n (%)			0.01			0.07
Occasional	926 (72.91)	4790 (76.26)		392 (73.27)	2061 (76.90)	
Regular	344 (27.09)	1491 (23.74)		143 (26.73)	619 (23.10)	
Moderate-to-intensity physical activity[b], n (%)			0.00			0.18
Never	225 (21.99)	806(17.31)		90 (21.23)	362 (18.43)	
Regular	798 (78.01)	3849(82.69)		334 (78.77)	1602 (81.57)	

Notes: [a]The dietary factors including the intake of fatty meat, fried food, and pickled food were categorized as occasional (<3 days per week) or regular (3–7 days per week). [b]Moderate-to-intensity physical activity was classified as never or regular (at least 1 day per week).
Abbreviations: SD, standard deviation; BMI, body mass index.

Discussion

Using data from a population-based screening cohort, this nested case–control study was performed to build a risk prediction model so as to identify average residents at high risk of early colorectal neoplasm. The model incorporating age, sex, family history, BMI, smoking status, alcohol, and regular moderate-to-intensity physical activity showed good reliability and discriminatory accuracy. AUC estimates were acceptable-to-good in both the training dataset and the validation dataset (AUC in the training dataset, 0.624; 95% CI, 0.604–0.643; AUC in the validation dataset, 0.630; 95% CI, 0.604–0.655). This study was novel in building a risk prediction model for early colorectal neoplasm among Chinese population based on lifestyle and risk factors for pre-selection in screening.

This risk stratification of residents to be screened has potential advantages. Several countries and cities have organized cancer screening programs with age as the

Table 3 Univariate and Multivariate Predictors of Early Colorectal Neoplasm in the Training Set

	Unadjusted		Adjusted		
	OR (95% CI)	**P value**	**β Coefficient**	**OR (95% CI)**	**P value**
Age (year)		0.00			0.00
50-	1		0.00	1	
60-	1.32 (1.16, 1.49)	0.00	0.37	1.45 (1.25, 1.69)	0.00
70-	1.15 (0.93, 1.42)	0.21	0.31	1.37 (1.07, 1.75)	0.01
Sex					
Male	1.84 (1.63, 2.08)	0.00	0.30	1.35 (1.09, 1.67)	0.01
Female	1		0.00	1	
BMI (kg/m^2)		0.00			0.01
18.5 ≤ BMI < 24	1		0.00	1	
BMI <18.5	0.97 (0.62, 1.52)	0.89	0.09	1.10 (0.69, 1.76)	0.68
24 ≤ BMI < 28	1.13 (0.98, 1.31)	0.10	0.12	1.13 (0.97, 1.31)	0.13
BMI ≥ 28	1.55 (1.23, 1.95)	0.00	0.41	1.51 (1.18, 1.92)	0.00
Family history					0.00
No	1		0.00	1	
Yes	2.76 (2.12, 3.59)	0.00	1.02	2.78 (2.08, 3.71)	
Smoking					0.02
No	1		0.00	1	
Current/ever	1.77 (1.57, 2.00)	0.00	0.25	1.28 (1.04, 1.58)	
Alcohol					0.03
No	1		0.00	1	
Current/ever	1.78 (1.56, 2.03)	0.00	0.20	1.23 (1.02, 1.48)	
Intake of fatty food[a]					0.65
Occasional	1		0.00	1	
Regular	1.15 (1.00, 1.33)	0.04	−0.04	0.96 (0.79, 1.15)	
Intake of fried food[a]					0.12
Occasional	1		0.00	1	
Regular	1.30 (1.05, 1.63)	0.02	0.22	1.25 (0.94, 1.65)	
Intake of pickled food[a]					0.27
Occasional	1		0.00	1	
Regular	1.19 (1.04, 1.37)	0.01	0.09	1.10 (0.93, 1.31)	
Moderate-to-intensity physical activity[b]					0.01
Never	1		0.00	1	
Regular	0.74 (0.63, 0.88)	0.00	− 0.24	0.79 (0.66, 0.94)	

Notes: [a]The dietary factors including intake of fatty meat, fried food and pickled food were categorized as occasional (<3 days per week) or regular (3–7 days per week). [b]Moderate-to-intensity physical activity was classified as never or regular (at least 1 day per week).

Abbreviations: BMI, body mass index; OR, odds ratio; 95% CI, 95% confidence interval..

only risk determinant. Yet cancer risk is affected by not only age but also other known variables. Based on the risk factors, multivariable cancer risk models could offer a better opportunity for risk assessment and stratification. Risk stratification could make a pre-selection for screening and thus help accurately determine the following step. Those identified with increased risks should be particularly motivated to receive screening. Compared with the traditional strategy, screening after risk stratification may significantly increase the participation and positive rate of colonoscopy, which is especially beneficial to resource-limited countries.

A number of risk prediction models have been developed to stratify high-risk people from average ones with

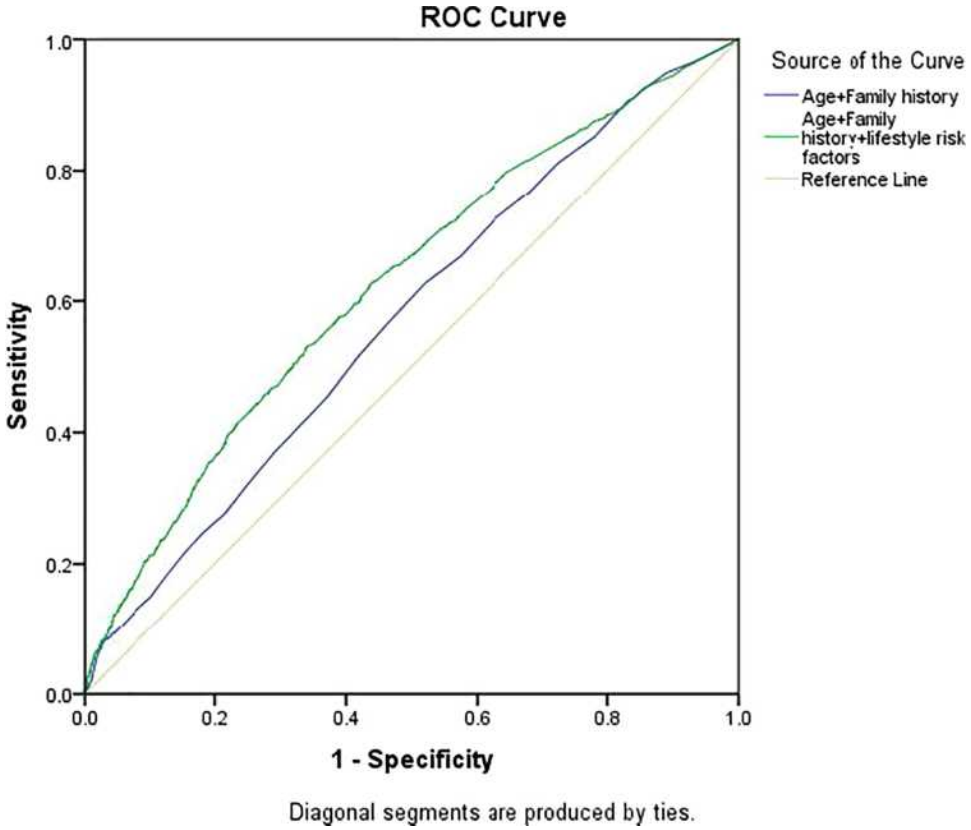

Figure 2 ROC curves for early colorectal neoplasm risk predictions. The yellow solid line is reference line. The solid line with blue is for risk model with only age and family history of first-degree relatives with CRC. The green line is for risk model with age, family history of first-degree relatives with CRC and lifestyle risk factors. The horizontal line corresponds to 1-specificity. The vertical line corresponds to sensitivity.

different risk factors for different people,[16–34] mostly among Asians, US adults and Europeans in high-resource countries (Table 5). Three previous reviews of the prediction models have also been published.[35–37] The proposed model showed acceptable-to-good discrimination in both training and validation datasets, with AUC estimates similar to those of previous studies.[18,19,22,23,26,29,30,32] To note, this model displayed lower discrimination compared with the ones proposed by Stageman et al.[33] Li et al[27] Aniwan et al[23] and Cooper et al[25] mainly because they incorporated qualitative FIT in the model and the model proposed in the present study was a pre-selection for high-risk people to participate in screening. Three models developed in China had higher AUC estimates: Cai et al[20] developed a model for ACN in 2012 based on age, sex, smoking, diabetes mellitus, green vegetables, pickled food, fried food, and white meat (AUC, 0.74); Chen et al[31] developed one model for ACN with multiple variables including age, sex, history of coronary heart disease, egg intake, defaecation frequency in 2014 (AUC, 0.75); Li et al[27] included age, sex, education level, occupations, diarrhea, constipation, colon mucosa and bleeding, gallbladder disease, a stressful life event, family history and a qualitative FIT for CRC model (AUC, 0.84). However, the former two models were built from a hospital-based population, and the last one considered the history of clinical symptoms; hence, the selection bias might not be corrected. The population-based risk prediction model with acceptable-to-good discrimination in this study may be of value for pre-screening in low-resource countries.

Table 4 Sensitivity and Specificity of Early Colorectal Neoplasm Risk-Based Model at Different Predicted Risk Cutoff Values

Predicted Risk Cutoff Values (%)	Sensitivity (%)	Specificity (%)	Youden's Index (%)
10	95.27	8.73	4.00
15	75.21	39.97	15.17
17	63.99	53.82	17.81
21	43.62	73.66	17.28
25	26.23	86.19	12.43
30	21.20	90.00	7.56
35	6.48	95.62	4.54
40	4.42	98.05	3.45

Table 5 Summary of Risk Prediction Models for CRC and ACN

Author (Year)	Country	Outcome	Risk Factors	Data Source (Sample Size)	AUC
Betes 2003[29]	Spain	ACN	Age, gender, BMI	Hospital-based (2210)	0.67
Cai 2012[20]	China	ACN +CRC	Age, gender, smoking, diabetes mellitus, green vegetables, pickled food, fried food, white meat	Hospital-based (7541)	0.74
Chen 2014[31]	China	ACN	Age, gender, history of coronary heart disease, egg intake, defaecation frequency	Hospital-based (905)	0.75
Li 2018[27]	China	CRC	Age, sex, education level, occupations, diarrhea, constipation, colon mucosa and bleeding, gallbladder disease, a stressful life event, family history and a qualitative FIT	Population based (891,199) considers the history of clinical symptoms and signs of bowel disease together with the FIT results	0.84
Kaminski 2014[32]	Poland	ACN	Age, gender, BMI, smoking, number and age affected of first degree relatives with CRC	Cross-sectional analysis (35,918)	0.64
Stegeman 2014[33]	Netherlands	ACN	Age, smoking, first degree relative with CRC, FIT, calcium intake	Population-based (6600)	0.76
Tao 2014[22]	Germany	ACN	Age, gender, smoking, first-degree relative with CRC, alcohol, previous polyp, red meat consumption, NSAIDS, previous colonoscopy	Population-based (7891)	0.67
Tao 2014[22]	Germany	CRC	Age, gender, smoking, first-degree relative with CRC, alcohol, previous polyp, red meat consumption, NSAIDS, previous colonoscopy	Population-based (7891)	0.71
Yeoh 2011[30]	Asia	ACN	Age, gender, smoking, first degree relative with CRC	Hospital-based (2752)	0.66
Driver 2007[16]	USA	CRC	Age, BMI, history of smoking, weekly or daily alcohol use	Population-based (21,581 male)	0.7
Ma 2008[17]	Japan	CRC	Age, BMI, smoking, physical activity, alcohol	Population-based (28,115 male)	0.7
Taylor 2011[18]	USA	CRC	Age, first, second and third degree relatives with CRC	Population-based (431,153)	0.67
Yarnall 2013[19]	UK	CRC	BMI, smoking, alcohol, fibre intake, red meat intake, physical activity, 14 SNPs	META	0.63
Auge 2014[34]	Spain	ACN	Age, sex and categorized FIT results	Retrospective (3109)	0.67
Cooper 2017[25]	UK	CRC+ ACN	Age, sex, first time invite, previous non-responder, previous responder and quantitative FIT result	Population based (40,930)	0.68
Aniwan 2015[23]	Thailand	CRC	Age, sex, family history and smoking status as APCS risk score and a qualitative FIT	Hospital-based (948)	0.85

(*Continued*)

Table 5 (Continued).

Author (Year)	Country	Outcome	Risk Factors	Data Source (Sample Size)	AUC
Aniwan 2015[23]	Thailand	ACN	Age, sex, family history and smoking status as APCS risk score and a qualitative FIT	Hospital-based (948)	0.67
Dunlop 2013[21]		CRC	Family history of CRC and 10 common genetic variants	Population-based (42,103)	0.56
Hsu 2015[24]	USA and Germany	CRC	Family history of CRC and 27 common genetic variants, endoscopy history	Population-based (12,000)	0.59 for men; 0.56 for women
Ibáñez-Sanz 2017[26]	Spain	CRC	Family history of CRC,alcohol consumption, obesity, physical activity, red meat and vegetable consumption, and nonsteroidal anti-inflammatory drug use, 21 SNPs	Population-based (4080)	0.63
Joen 2018[28]	USA	CRC	Family history of CRC, 19 lifestyle and environmental factors, 63 CRC-associated SNPs	Population-based (20,338)	0.63 for men; 0.62 for women

Abbreviations: CRC, colorectal cancer; ACN, advanced colorectal neoplasm; FIT, faecal immunochemical test; NSAIDS, nonsteroidal anti-inflammatory drug use; BMI, body mass index; SNP, single nucleotide polymorphism; APCS, Asia-Pacific Colorectal Screening Scoring System.

The multivariable analysis showed that older age; male sex; obesity; smoking habit; alcohol drinking status; higher consumption of fatty food, pickled food, or fried food; and lower physical activity were predictors of early colorectal neoplasm, which was consistent with the results from previous investigations.[38–42] Age is a major determinant of CRC risk; therefore, cancer screening is usually not recommended to young people. The finding that men were at increased risk of early colorectal neoplasm was consistent with other finding on different populations worldwide. Sufficient evidence indicated that a family history of CRC was significantly associated with a higher risk of colorectal neoplasm.[21–24,26,33] The present study also demonstrated a close association between positive family history of CRC and early colorectal neoplasm (OR = 2.78, 95% CI: 2.08–3.71). It has been reported widely that lifestyle is a well-established risk factor for CRC and associated with the occurrence of adenomas.[20,38] The result also showed that obesity, smoking, and alcohol drinking were associated with an increased risk of early colorectal neoplasm, while regular moderate-to-intensity physical activity was a protective factor, consistent with previous findings.[39,42–45]

The proposed model was reliable and practical. On the one hand, the risk-based model could provide accurate information about the personal risks of colorectal neoplasm. On the other hand, the perceptions of personal risks might be translated into behavior and lifestyle changes, which would potentially reduce personal risks or prevent cancer. As advocated by Stegeman I,[46] screening should be performed not only to find early cancer cases but also to prevent cancer by further interfering with risk factors. Besides, given the race and ethnicity differences in the CRC prevalence, the proposed model with specific Chinese characteristics could be more practical. When applying the risk model to screening practice, not only clinicians can tailor screening based on a person's predicted risk of colorectal neoplasm but also average people could have more chance to gain expected health benefits, such as life expectancy of screens, preventive risk factors, healthy lifestyle and disease. Hence, it was believed that the proposed model was not just a risk-based model but also a benefit-based model.

This study had several strengths and limitations. The major limitation of the study was the lack of external validation. Hence, further external large-population-based validation studies are required for model discrimination and implementation. Another limitation was that the data was from the screening cohort and not every participant received colonoscopy, which might have precluded the outcome evaluation and led to misclassification bias. However, the control group might have

probably more cases of early colorectal neoplasm undetected, thus diluting the association in the model and making the results relatively conservative. Moreover, the physical measurement, such as weight, in the present study was performed only once at the baseline, leading to difficulty in assessing the impact of possible change on the result. Besides, the self-report lifestyle habits would give rise to information bias. The most important strength of the study was that a population-based nested case–control study with data collected from a large-population community-based cohort could minimize the possibility that the associations in the study for building the model were due to early disease effects compared with normal retrospective case–control studies and maximize the stability of the model. In addition, the information was obtained by experienced physicians through face-to-face interview, and the physical indicators such as height and weight were measured using unified instruments, minimizing the detection and information biases.

In conclusion, a risk determination model with a better discriminator power was developed in this study for early colorectal neoplasm that might optimize the screening strategy by identifying residents at increased risks. This risk-based model could improve the pre-selection for screening and contribute a lot to efficient population-based screening in low-resource countries. People at high risk might be entitled to change their behaviors and lifestyles so that they would achieve more health benefits. Artificial intelligence (AI) models will change current CRC screening strategy in the near future when AI technology is increasingly used in CRC detection.[47] Previously hampered by poor participation and poor performance of community endoscopists, the future AI-based CRC model will make screening universally feasible and reliable.

Acknowledgments

The authors thank all the participants, the primary care physicians and their institutions for their contributions to this study.

Funding

This study was supported by the Grant of National Natural Science Foundation of China (No. 81871958), the Grant of Shanghai Municipal Health Commission (No. 2020JZX0206), and the Fudan University Shanghai Cancer Center Institutional Fund (No. YJRC1701).

Disclosure

The authors declared no conflict of interest for this work.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394–424. doi:10.3322/caac.21492.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin*. 2018;68(1):7–30. doi:10.3322/caac.21442.
3. Chen W, Zheng R, Zhang S, et al. Cancer incidence and mortality in China, 2013. *Cancer Lett*. 2017;401:63–71. doi:10.1016/j.canlet.2017.04.024.
4. Zhang L, Cao F, Zhang GY, et al. Trends in and predictions of colorectal cancer incidence and mortality in China from 1990 to 2025. *Front Oncol*. 2019;9:98. doi:10.3389/fonc.2019.00098
5. Lieberman D, Ladabaum U, Cruz-Correa M, et al. Screening for colorectal cancer and evolving issues for physicians and patients: a review. *JAMA*. 2016;316(20):2035–2045. doi:10.1001/jama.2016.17418
6. Siegel RL, Miller KD, Fedewa SA, et al. Colorectal cancer statistics, 2017. *CA Cancer J Clin*. 2017;67(3):177–193. doi:10.3322/caac.21395
7. Dekker E, Rex DK. Advances in CRC prevention: screening and surveillance. *Gastroenterology*. 2018;154(7):1970–1984. doi:10.1053/j.gastro.2018.01.069.
8. Gong Y, Bao P, Zhong W, et al. The implementation and first-round results of a community-based colorectal cancer screening program in Shanghai, China. *Oncologist*. 2018;23(8):928–935. doi:10.1634/theoncologist.2017-0451
9. Klabunde Carrie N, Joseph Djenaba A, King Jessica B, et al. Vital signs: colorectal cancer screening test use – United States, 2012. *MMWR Morb Mortal Wkly Rep*. 2013;62:881–888.
10. Bretthauer M, Kaminski MF, Løberg M, et al. Population-based colonoscopy screening for colorectal cancer: a randomized clinical trial. *JAMA Intern Med*. 2016;176(7):894–902. doi:10.1001/jamainternmed.2016.0960.
11. Chen H, Li N, Ren J; group of Cancer Screening Program in Urban China (CanSPUC), et al. Participation and yield of a population-based colorectal cancer screening programme in China. *Gut*. 2019;68(8):1450–1457. doi:10.1136/gutjnl-2018-317124.
12. Klabunde C, Blom J, Bulliard JL, et al. Participation rates for organized colorectal cancer screening programmes: an international comparison. *J Med Screen*. 2015;22(3):119–126. doi:10.1177/0969141315584694
13. An J, Sun Z, Li Y, Lu X, Ji G. Literature review and consideration of factors influencing colorectal cancer screening in China. *Chin Gen Prac*. 2020;23(23):2877–2882. doi:10.12114/j.issn.1007-9572.2020.00.160
14. Yuan P, Gu J. Meta-analysis of the compliance of colorectal cancer screening in China, 2006～2015. *China Cancer*. 2017;26(4):241-248. doi:10.11735/j.issn.1004-0242.2017.04.A001
15. Li XP, Qian MC, Zhao GM, et al. The performance of a community-based colorectal cancer screening program: evidence from Shanghai Pudong New Area, China. *Prev Med*. 2019;118:243–250. doi:10.1016/j.ypmed.2018.11.002
16. Driver JA, Gaziano JM, Gelber RP, et al. Development of a risk score for colorectal cancer in men. *Am J Med*. 2007;120(3):257–263. doi:10.1016/j.amjmed.2006.05.055.
17. Ma E, Sasazuki S, Iwasaki M, Sawada N, Inoue M, Tsugane S. Japan Public Health Center-based Prospective Study Group. 10-Year risk of colorectal cancer: development and validation of a prediction model in middle-aged Japanese men. *Cancer Epidemiol*. 2010;34(5):534–541. doi:10.1016/j.canep.2010.04.021
18. Taylor DP, Stoddard GJ, Burt RW, et al. How well does family history predict who will get colorectal cancer? Implications for cancer screening and counseling. *Genet Med*. 2011;13(5):385–391. doi:10.1097/GIM.0b013e3182064384

19. Yarnall JM, Crouch DJM, Lewis CM. Incorporating non-genetic risk factors and behavioural modifications into risk prediction models for colorectal cancer. *Cancer Epidemiol*. 2013;37:324–329. doi:10.1016/j.canep.2012.12.008
20. Cai QC, Yu ED, Xiao Y, et al. Derivation and validation of a prediction rule for estimating advanced colorectal neoplasm risk in average-risk Chinese. *Am J Epidemiol*. 2012;175(6):584–593. doi:10.1093/aje/kwr337
21. Dunlop MG, Tenesa A, Farrington SM, et al. Cumulative impact of common genetic variants and other risk factors on colorectal cancer risk in 42,103 individuals. *Gut*. 2013;62(6):871–881. doi:10.1136/gutjnl-2011-300537
22. Tao S, Hoffmeister M, Brenner H. Development and validation of a scoring system to identify individuals at high risk for advanced colorectal neoplasms who should undergo colonoscopy screening. *Clin Gastroenterol Hepatol*. 2014;12(3):478–485. doi:10.1016/j.cgh.2013.08.042.
23. Aniwan S, Rerknimitr R, Kongkam P, et al. A combination of clinical risk stratification and fecal immunochemical test results to prioritize colonoscopy screening in asymptomatic participants. *Gastrointest Endosc*. 2015;81(3):719–727. doi:10.1016/j.gie.2014.11.035
24. Hsu L, Jeon J, Brenner H, et al. A model to determine colorectal cancer risk using common genetic susceptibility loci. *Gastroenterology*. 2015;148(7):1330–1339.e14. doi:10.1053/j.gastro.2015.02.010
25. Cooper JA, Parsons N, Stinton C, et al. Risk-adjusted colorectal cancer screening using the FIT and routine screening data: development of a risk prediction model. *Br J Cancer*. 2018;118(2):285–293. doi:10.1038/bjc.2017.375
26. Ibáñez -sanz G, Diez-Villanueva A, Alonso MH, et al. Risk model for colorectal cancer in Spanish population using environmental and genetic factors: results from the MCC-Spain study. *Sci Rep*. 2017;7:43263. doi:10.1038/srep43263.
27. Li W, Zhao LZ, Ma DW, et al. Predicting the risk for colorectal cancer with personal characteristics and fecal immunochemical test. *Medicine (Baltimore)*. 2018;97:e0529. doi:10.1097/MD.0000000000010529
28. Jeon J, Du MM, Schoen RE, et al. Determining risk of colorectal cancer and starting age of screening based on lifestyle, environmental, and genetic factors. *Gastroenterology*. 2018;154(8):2152–2164. doi:10.1053/j.gastro.2018.02.021
29. Betés M, Muñoz-Navas MA, Duque JM, et al. Use of colonoscopy as a primary screening test for colorectal cancer in average risk people. *Am J Gastroenterol*. 2003;98(12):2648–2654. doi:10.1111/j.1572-0241.2003.08771.x
30. Yeoh KG, Ho KY, Chiu HM, et al. The Asia-Pacific colorectal screening score a validated tool that stratifies risk for colorectal advanced neoplasia in asymptomatic Asian subjects. *Gut*. 2011;60(9):1236–1241. doi:10.1136/gut.2010.221168
31. Chen G, Mao B, Pan Q, Liu Q, Xu X, Ning Y. Prediction rule for estimating advanced colorectal neoplasm risk in average-risk populations in southern Jiangsu Province. *Chin J Cancer Res*. 2014;26(1):4–11. doi:10.3978/j.issn.1000-9604.2014.02.03
32. Kaminski MF, Polkowski M, Kraszewska E, Rupinski M, Butruk E, Regula J. A score to estimate the likelihood of detecting advanced colorectal neoplasia at colonoscopy. *Gut*. 2014;63(7):1112–1119. doi:10.1136/gutjnl-2013-304965
33. Stegeman I, de Wijkerslooth TR, Stoop EM, et al. Combining risk factors with faecal immunochemical test outcome for selecting CRC screenees for colonoscopy. *Gut*. 2014;63(3):466–471. doi:10.1136/gutjnl-2013-305013
34. Auge JM, Pellise M, Escudero JM, et al. Risk stratification for advanced colorectal neoplasia according to fecal hemoglobin concentration in a colorectal cancer screening program. *Gastroenterology*. 2014;147(3):628–636.e1. doi:10.1053/j.gastro.2014.06.008
35. Usher-Smith JA, Walter FM, Emery JD, et al. Risk prediction models for colorectal cancer: a systematic review. *Cancer Prev Res*. 2016;9(1):13–26. doi:10.1158/1940-6207.CAPR-15-0274
36. Win AK, Macinnis RJ, Hopper JL, Jenkins MA. Risk prediction models for colorectal cancer: a review. *Cancer Epidemiol Biomarkers Prev*. 2012;21:398–410. doi:10.1158/1055-9965.EPI-11-0771
37. Ma GK, Ladabaum U. Personalizing colorectal cancer screening: a systematic review of models to predict risk of colorectal neoplasia. *Clin Gastroenterol Hepatol*. 2014;12:1624–34.e1. doi:10.1016/j.cgh.2014.01.042
38. Larsen IK, Grotmol T, Almendingen K, Hoff G. Lifestyle as a predictor for colonic neoplasia in asymptomatic individuals. *BMC Gastroenterol*. 2006;6(1):5. doi:10.1186/1471-230X-6-5
39. Moskal A, Norat T, Ferrari P, Riboli E. Alcohol intake and colorectal cancer risk: a dose–response meta-analysis of published cohort studies. *Int J Cancer*. 2007;120(3):664–671. doi:10.1002/ijc.22299
40. Lieberman DA, Prindiville S, Weiss DG, Willett W, VA Cooperative Study Group 380. Risk factors for advanced colonic neoplasia and hyperplastic polyps in asymptomatic individuals. *JAMA*. 2003;290(22):2959–2967. doi:10.1001/jama.290.22.2959
41. Roos VH, Mangas-Sanjuan C, Rodriguez-Girondo M, et al. Effects of family history on relative and absolute risks for colorectal cancer: a systematic review and meta-analysis. *Clin Gastroenterol Hepatol*. 2019;17(13):2657–2667.e9. doi:10.1016/j.cgh.2019.09.007
42. Choi YJ, Lee DH, Han KD, et al. The relationship between drinking alcohol and esophageal, gastric or colorectal cancer: a nationwide population-based cohort study of South Korea. *PLoS One*. 2017;12(10):e0185778. doi:10.1371/journal.pone.0185778
43. Botteri E, Iodice S, Bagnardi V, Raimondi S, Lowenfels AB, Maisonneuve P. Smoking and colorectal cancer: a meta-analysis. *JAMA*. 2008;300(23):2765–2778. doi:10.1001/jama.2008.839
44. Bardou M, Barkun AN, Martel M. Obesity and colorectal cancer. *Gut*. 2013;62(6):933–947. doi:10.1136/gutjnl-2013-304701
45. Soltani G, Poursheikhani A, Yassi M, Hayatbakhsh A, Kerachian M, Kerachian MA. Obesity, diabetes and the risk of colorectal adenoma and cancer. *BMC Endocr Disord*. 2019;19(1):113. doi:10.1186/s12902-019-0444-6
46. Stegeman I, Bossuyt PM. Cancer risk models and preselection for screening. *Cancer Epidemiol*. 2012;36(5):461–469. doi:10.1016/j.canep.2012.06.010
47. Goyal H, Mann R, Gandhi Z, et al. Scope of artificial intelligence in screening and diagnosis of colorectal cancer. *J Clin Med*. 2020;9(10):3313. doi:10.3390/jcm9103313