Cancer Letters xxx (xxxx) 217057

Contents lists available at ScienceDirect

Cancer Letters

journal homepage: www.elsevier.com/locate/canlet

ELSEVIER

Original Articles

Development and evaluation of a risk prediction tool for risk-adapted screening of colorectal cancer in China

Dong Hang[a,b,1], Dianjianyi Sun[c,d,e,1], Lingbin Du[f,g,1], Jianv Huang[a], Jiacong Li[a], Chen Zhu[a,f,g], Le Wang[f,g], Jingjing He[a], Xia Zhu[a], Meng Zhu[a,b], Ci Song[a,b], Juncheng Dai[a,b], Canqing Yu[c,d,e], Zekuan Xu[h], Ni Li[i], Hongxia Ma[a,b], Guangfu Jin[a,b], Ling Yang[j,k], Yiping Chen[j,k], Huaidong Du[j,k], Xiangdong Cheng[f,g], Zhengming Chen[k], Jun Lv[c,d,e,1], Zhibin Hu[a,b,*], Liming Li[c,d,e,**], Hongbing Shen[a,b,l,***], China Kadoorie Biobank Collaborative Group

[a] Department of Epidemiology, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, China
[b] Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Medicine, China International Cooperation Center for Environment and Human Health, Nanjing Medical University, Nanjing, China
[c] Department of Epidemiology and Biostatistics, School of Public Health, Peking University Health Science Center, Beijing, China
[d] Peking University Center for Public Health and Epidemic Preparedness and Response, Beijing, China
[e] Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Beijing, China
[f] Department of Cancer Prevention, The Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital), Hangzhou, China
[g] Institute of Basic Medicine and Cancer (IBMC), Chinese Academy of Sciences, Hangzhou, China
[h] Department of General Surgery, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China
[i] Office of Cancer Screening, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
[j] Medical Research Council Population Health Research Unit, University of Oxford, Oxford, United Kingdom
[k] Clinical Trial Service Unit & Epidemiological Studies Unit (CTSU), Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom
[l] Research Units of Cohort Study on Cardiovascular Diseases and Cancers, Chinese Academy of Medical Sciences, Beijing, China

ARTICLE INFO

Keywords:
Colorectal cancer
Risk prediction
Screening
Cohort study

ABSTRACT

Risk prediction tools for colorectal cancer (CRC) have potential to improve the efficiency of population-based screening by facilitating risk-adapted strategies. However, such an applicable tool has yet to be established in the Chinese population. In this study, a risk score was created using data from the China Kadoorie Biobank (CKB), a nationwide cohort study of 409,854 eligible participants. Diagnostic performance of the risk score was evaluated in an independent CRC screening programme, which included 91,575 participants who accepted colonoscopy at designed hospitals in Zhejiang Province, China. Over a median follow-up of 11.1 years, 3136 CRC cases were documented in the CKB. A risk score was created based on nine questionnaire-derived variables, showing moderate discrimination for 10-year CRC risk (C-statistic = 0.68, 95 % CI: 0.67–0.69). In the CRC screening programme, the detection rates of CRC were 0.25 %, 0.82 %, and 1.93 % in low-risk (score <6), intermediate-risk (score: 6–19), and high-risk (score >19) groups, respectively. The newly developed score exhibited a C-statistic of 0.65 (95 % CI: 0.63–0.66), surpassing the widely adopted tools such as the Asia-Pacific Colorectal Screening (APCS), modified APCS, and Korean Colorectal Screening scores (all C-statistics = 0.60). In conclusion, we developed a novel risk prediction tool that is useful to identify individuals at high risk of CRC. A user-friendly online calculator was also constructed to encourage broader adoption of the tool.

* Corresponding author. Department of Epidemiology, School of Public Health, Nanjing Medical University, 101 Longmian Avenue, Nanjing, 211166, China.
** Corresponding author. Department of Epidemiology and Biostatistics, School of Public Health, Peking University Health Science Center, 38 Xueyuan Road, Beijing, 100191, China.
*** Corresponding author. Department of Epidemiology, School of Public Health, Nanjing Medical University, 101 Longmian Avenue, Nanjing, 211166, China.
E-mail addresses: zhibin_hu@njmu.edu.cn (Z. Hu), lmlee@bjmu.edu.cn (L. Li), hbshen@njmu.edu.cn (H. Shen).

[1] Co-first authors, contributed equally.

[2] The members of steering committee and collaborative group are listed in the online-only supplementary material.

https://doi.org/10.1016/j.canlet.2024.217057
Received 1 March 2024; Received in revised form 3 June 2024; Accepted 10 June 2024

1. Introduction

Colorectal cancer (CRC) ranks as the third most common cancer and the second leading cause of cancer death worldwide, causing an estimated 1.9 million new cases and 0.9 million deaths in 2020 [1]. While CRC incidence and mortality have been gradually decreasing in highly developed countries through timely detection and treatment, a rapid upward trend has been seen in many low- and middle-income countries [1]. China, for instance, has recorded a significant surge in CRC cases, with average annual percent changes of 3.8 % in incidence and 1.2 % in mortality between 2000 and 2019 [2]. The trend could be primarily attributed to the population aging, lifestyle changes, adoption of Western diets, and unmet screening programmes [3].

Although colonoscopy is regarded as the gold standard tool for CRC screening, its diagnostic yield in the general population is not optimal and remains a major challenge. A previous large-scale screening programme in China showed that less than 4 % of adults who underwent colonoscopy were diagnosed with advanced neoplasia, which means that the majority of participants received an invasive procedure without significant direct benefit [4]. Moreover, given the huge population, limited healthcare resources, and low adherence rates, a one-size-fits-all screening service is not suitable in China [5]. Therefore, preselection of high-risk individuals for CRC screening has been suggested as a potentially cost-effective approach to reduce the cancer burden.

Currently, various CRC risk prediction models have been developed for use in the general population, which typically included age, sex, family history, and body mass index (BMI) [6]. However, the performance of these models in identifying individuals at high risk for CRC varied greatly, with C-statistics ranging from 0.61 to 0.73 [7]. The variation is likely due to the differences in study design, population characteristics, sample size, and selected variables in the models [7]. Although a few CRC risk prediction models have been built in the Chinese population, most relevant studies have limitations such as cross-sectional or case-control designs with potential selection bias and reverse causality [8–10], small sample sizes (n < 1000) with insufficient statistical power [9], or lack of external validation necessary for translation into practice [9–11]. Given apparent heterogeneity of lifestyle and dietary habits in different populations, an applicable risk prediction model specifically developed for the Chinese population is urgently needed.

In the current study, we created a novel risk prediction tool for CRC based on the China Kadoorie Biobank (CKB), a nationwide prospective cohort. We further evaluated its applicability in a real-world, large-scale colonoscopy screening programme. We also constructed a user-friendly online calculator named Colorectal Cancer Risk Evaluation Score (CORES), which is readily available to the general public and aims to encourage broader adoption of the tool (http://ccra.njmu.edu.cn/cores/web).

2. Materials and methods

2.1. Study design and population

The current study was conducted in two stages, as depicted by the flowchart in Supplemental Fig. 1. First, we developed a risk prediction model for CRC using data from the CKB, which recruited 512,725 participants aged 30–79 years from 10 regions of China between June 2004 and July 2008. At baseline, all participants completed an interviewer-administered electronic questionnaire and underwent physical examinations. Details of the CKB study design, survey methods, and follow-up have been described previously [12,13]. We excluded participants who were <40 or >75 years old (n = 81,092), had a history of cancer prior to baseline (n = 2441), or had missing data on the studied factors (n = 19,338), leaving a final sample size of 409,854 individuals for constructing the risk prediction model.

Second, we evaluated the applicability of the model in a large-scale CRC screening programme in Zhejiang Province, which is an ongoing community-based screening initiated since June 2020. Permanent residents aged 50–74 years from 11 cities were invited to participate in the programme. To encourage participation, individuals who were sightly outside the target age range were not turned away. Exclusion criteria included a previous history of CRC, colon resection, receipt of cancer therapy, and health problems that made colonoscopy unsuitable. The screening procedure included fecal immunochemical testing (FIT) and subsequent colonoscopy for individuals with FIT positivity. Community healthcare staff distributed two FIT packages to each participant and provided instructions on completing the tests at home within an interval of seven days. The threshold for FIT positivity was set at 20 μg Hb/g feces (equal to 100 ng Hb/mL), as recommended by the manufacturer. An individual was considered as FIT-positive if either one or both of the tests were positive. From June to December 2020, a total of 2,051,064 adults completed FIT. Among 241,051 participants who were FIT-positive, 97,073 (40.3 %) underwent colonoscopy at designated hospitals. According to the measure of standardized mean differences [14], there were no statistically significant differences in the distribution of demographic characteristics and lifestyle factors between the FIT-positive and total participants (Supplemental Table 1), and also no statistically significant differences between attendees and non-attendees of the colonoscopic examination among the FIT-positive individuals (Supplemental Table 2). We excluded those with prior cancer history (n = 2113), missing data on the studied factors (n = 3096), or no pathologic diagnosis (n = 289), resulting in 91,575 participants for the model evaluation.

The CKB was approved by the Ethical Review Committee of the Chinese Center for Disease Control and Prevention (Beijing, China: 005/2004) and the Oxford Tropical Research Ethics Committee at University of Oxford (UK: 025–04). The Zhejiang CRC screening programme was approved by the Ethics Committee and Institutional Review Board of Zhejiang Cancer Hospital (Zhejiang, China: IRB-2023-464). Written informed consent was obtained from all participants before enrollment.

2.2. Assessment of risk factors

Self-reported information on demographic characteristics, lifestyle factors, dietary habits, and medical history was collected using similar structured questionnaires in the CKB and Zhejiang screening programme. Body mass index (BMI) was calculated by dividing weight in kilograms by height in meters squared. The examined factors used to develop the prediction model were potential CRC risk factors [2,15,16], which included age, sex, education level, smoking status and pack-years, alcohol drinking, dietary factors (red meat, poultry, fish or sea food, fresh fruit and vegetables, and preserved vegetables), physical activity, BMI, prevalent diabetes, history of peptic ulcer disease, gallstone disease, and family history of cancer in first-degree relatives.

In the model development, age was divided into 5-year interval groups. Education was classified as illiterate or primary school, middle school, high school and above. Tobacco smoking was categorized as never or light smoking (<30 pack-years) and heavy smoking (≥30 pack-years) [17]. Alcohol consumption was categorized as never or occasional drinking (<3 days per week) and frequent drinking (≥3 days per week) [18]. Frequent intake of red meat, poultry, fish or sea food, fresh fruit and vegetables, or preserved vegetables during the previous 12 months was defined as ≥4 days per week, otherwise was occasional [19]. Total physical activity was classified as active and inactive intensity according to the median value of metabolic equivalent-hours/day (≥17.5 and < 17.5 MET-h/d). BMI was categorized into ≥28 and < 28 kg/m^2, based on the cut-off point for obesity in Chinese adults [20].

2.3. Ascertainment of outcome

CRC cases were recorded according to the 10th Revision of the International Classification of Diseases, including cancers occurring at the colon (C18), rectosigmoid junction (C19), and rectum (C20). In the CKB, incident CRC was identified through linkage with established chronic disease registries, national health insurance claim databases, and the disease surveillance points system [13].

In the Zhejiang screening programme, all colonoscopies and therapies were performed by experienced endoscopists following standard procedures. Abnormal findings during colonoscopy were confirmed by pathological examinations. CRC was diagnosed according to the Chinese Protocol of Diagnosis and Treatment of Colorectal Cancer (2017 edition), which was the primary outcome of the current study. Advanced adenoma was defined as at least one adenoma of ≥10 mm in diameter or with advanced histology (tubulovillous/villous histological features or high-grade/severe dysplasia) [21,22]. When more than one lesion was identified, the diagnosis was made based on the most severe one.

2.4. Statistical analysis

Person-time for participants was calculated from baseline until the date of CRC diagnosis, death, loss to follow-up, or the end of the follow-up (December 31, 2017), whichever occurred first in the CKB. We used Cox regression models to examine the association between variables and CRC risk, computing hazards ratios (HRs) and 95 % confidence intervals (CIs). Schoenfeld residuals were applied to test the proportional hazards assumption [23]. To identify the best predictors among plausible predictive variables, we used multivariable Cox regression with stepwise selection of $P < 0.10$ for entry of variables and $P > 0.10$ for removal of variables. Regression coefficients from the final model were adopted to calculate a CRC risk prediction score. Specifically, the variable with the smallest regression coefficient in the model was assigned one point, while other variables were assigned points based on the ratios of their corresponding coefficients against the smallest coefficient. These points were rounded up to the nearest integer and summed up to create a risk prediction score for each participant [11,24].

Discrimination was assessed by the area under the receiver operating characteristic curve (AUC), also known as C-statistics, for 10-year risk model. Calibration was performed by plotting the mean observed probability (Kaplan-Meier survival function) against the mean predicted probability of CRC at 10 years within each decile of the score (Cox regression). In addition, linear regression was conducted, and the coefficient of determination (R^2) was used for quantitative calibration [25].

We further evaluated the ability of the risk score to stratify the population under different risk levels in the CKB. The absolute risk of CRC was projected at three time points (3, 5, and 10 years) by deciles of the risk score in the Cox regression model. In addition, participants were divided into low-risk (bottom 20 %), intermediate-risk (20%–80 %), and high-risk (top 20 %) groups based on their risk scores, and the corresponding 3-, 5-, and 10-year cumulative incidences were estimated.

To evaluate the usefulness of the risk score in a screening practice, we calculated the score for participants in the Zhejiang CRC screening programme, and classified them into ten groups using the decile cutoffs derived from the CKB. We compared the discrimination ability of the new score against commonly used models, including the Asia-Pacific Colorectal Screening (APCS) score [26], the modified APCS score [27], and the Korean Colorectal Screening (KCS) score [28], based on the DeLong test. Moreover, we calculated the net reclassification improvement (NRI) to evaluate the added predictivity of the new score [29].

In a sensitivity analysis to assess reverse causality, we repeated the variable-selection procedure in multivariable Cox regression after excluding 342 CRC cases diagnosed within the first two years of follow-up in the CKB. We also performed a competing risk model by considering death as a competing event, as the occurrence of death might prevent us from observing the development of CRC among those participants. Additionally, we conducted stratified analyses by sex and urban-rural areas in the CKB to determine whether the association between variables and CRC risk was consistent across strata. Statistical analyses were performed using R version 3.6.1 (R Core Team, Vienna, Austria). All tests were two-sided, and a P-value of < 0.05 was considered statistically significant.

3. Results

3.1. Study population

Among 409,854 eligible participants from the CKB, 3136 incident CRC cases were recorded over a median follow-up of 11.1 (interquartile range: 10.1–12.1) years. In the Zhejiang screening programme, 1185 (1.29 %) CRC and 10,313 (11.3 %) advanced adenoma cases were diagnosed. Table 1 shows the basic characteristics of study participants. The distributions of demographic and lifestyle characteristics were generally similar across the two cohorts, although the mean age of participants was higher in the Zhejiang screening programme (61.4 years) than in the CKB (54.3 years). Additionally, the proportion of frequent alcohol drinking was higher in the Zhejiang screening programme (22.4 %) than in the CKB (12.3 %).

3.2. Development of CRC risk prediction score

Nine variables were identified as independent predictors of CRC by the stepwise regression analysis. These including age, sex, BMI, pack-years of smoking, alcohol drinking, red meat intake, prevalent diabetes, history of gallstone disease, and family history of cancer in first-degree relatives (Table 2). The results remained basically unchanged when removing incident CRC cases within the first two years of follow-up (Supplemental Table 3) or considering death as a competing event in the competing risk model (Supplemental Table 4). In addition, the stratified analysis by sex showed that the associations were generally consistent between men and women, with slightly stronger associations noted for age and BMI in men compared to women (Supplemental Table 5). The stratified analysis by urban-rural areas also indicated that the associations were largely consistent across the subgroups, except that stronger associations were observed for age and sex in urban population compared to rural population, and a stronger association was found for red meat intake in rural population compared to urban population (Supplemental Table 6).

The calculated score ranged from 0 to 37, with higher scores observed in incident CRC cases (mean ± SD: 16.7 ± 7.2) compared to CRC-free individuals (12.3 ± 7.4) (Fig. 1A). Individuals aged 70–75 years constituted the most high-risk group (Supplemental Fig. 2). A significant increase in CRC incidence was noted with ascending score deciles, ranging from 25 to 186 per 100,000 person-years (P trend = 0.0001) (Supplemental Table 7). The C statistic of the score measuring discrimination for 10-year risk of CRC was 0.68 (95 % CI: 0.67–0.69), and the predicted risk by the score deciles was calibrated well with the observed 10-year CRC risk ($R^2 = 0.98$) (Fig. 1B and C).

3.3. Absolute risk of incident CRC

We estimated the absolute risk of incident CRC at 3, 5, and 10 years according to deciles of the risk score in the CKB (Supplemental Table 8). Comparing participants in the highest versus the lowest deciles, the HRs of CRC at 3, 5, and 10 years were 8.78 (95 % CI: 5.56–13.88), 9.51 (95 % CI: 6.86–13.17), and 7.00 (95 % CI: 5.73–8.54), respectively. We further categorized the participants into low-risk (bottom 20 % of the score: < 6), indeterminate-risk (20%–80 %: 6–19), and high-risk (top

Table 1
Basic characteristics of colorectal cancer cases in the CKB cohort and the Zhejiang CRC screening programme.

Variable[a]	CKB cohort			Zhejiang screening programme		
	Total (n = 409,854) No. (%)	CRC cases (n = 3136) No.	Incidence rate (per 100,000 person-years)	Total (n = 91,575) No. (%)	CRC cases (n = 1185) No.	Detection rate %
Age, years	54.28 ± 9.13	59.39 ± 8.97		61.44 ± 7.02	64.84 ± 6.22	
40-<45	81,867 (19.97)	235	25.76	742 (0.81)	2	0.27
45-<50	66,459 (16.22)	298	40.14	2236 (2.44)	6	0.27
50-<55	84,248 (20.56)	523	56.37	14,598 (15.94)	89	0.61
55-<60	66,368 (16.19)	561	78.26	18,418 (20.11)	135	0.73
60-<65	46,349 (11.31)	532	109.27	20,822 (22.74)	267	1.28
65-<70	37,980 (9.27)	548	144.20	21,547 (23.53)	375	1.74
70–75	26,583 (6.49)	439	177.43	13,212 (14.43)	311	2.35
Sex						
Female	241,835 (59.01)	1590	60.22	46,635 (50.93)	460	0.99
Male	168,019 (40.99)	1546	87.16	44,940 (49.07)	725	1.61
BMI, kg/m^2[b]	23.78 ± 3.39	24.22 ± 3.54		23.56 ± 2.99	23.74 ± 3.13	
<28	364,382 (88.91)	2681	68.35	84,820 (92.62)	1081	1.27
≥28	45,472 (11.09)	455	92.56	6755 (7.38)	104	1.54
Pack-years of smoking[c]						
<30	360,368 (87.93)	2559	65.57	79,343 (86.64)	993	1.25
≥30	49,486 (12.07)	577	112.93	12,232 (13.36)	192	1.57
Alcohol drinking[d]						
Occasional	359,606 (87.74)	2604	67.17	71,099 (77.64)	844	1.19
Frequent	50,248 (12.26)	532	99.06	20,476 (22.36)	341	1.67
Red meat intake[e]						
Occasional	217,470 (53.06)	1465	62.97	51,909 (56.68)	679	1.31
Frequent	192,384 (46.94)	1671	80.06	39,666 (43.32)	506	1.28
Prevalent diabetes						
No	382,649 (93.36)	2820	68.14	84,013 (91.74)	1037	1.23
Yes	27,205 (6.64)	316	114.90	7562 (8.26)	148	1.96
History of gallstone disease						
No	383,216 (93.50)	2881	69.80	88,446 (96.58)	1145	1.29
Yes	26,638 (6.50)	255	88.97	3129 (3.42)	40	1.28
Family history of cancer in first-degree relatives						
No	333,636 (81.40)	2490	69.40	77,403 (84.52)	1014	1.31
Yes	76,218 (18.60)	646	78.22	14,172 (15.48)	171	1.21

Abbreviations: BMI, body mass index; CKB, China Kadoorie Biobank; CRC, colorectal cancer.
[a] Continuous variables are presented as the mean ± standard deviation and categorical variables are presented as number (proportion).
[b] BMI was calculated as weight in kilograms divided by the square of height in meters.
[c] Pack-year was calculated by the product of the years of smoking (excluding years of quitting smoking) and the number of cigarette packs (the number of cigarettes divided by 20) smoked per day.
[d] Frequent alcohol drinking was defined as drinking ≥3 days per week, otherwise was occasional.
[e] Frequent intake of red meat was defined as eating red meat ≥4 days per week in the CKB and Zhejiang screening programme, otherwise was occasional. In the Changzhou cohort, frequent intake of red meat was defined as eating red meat ≥3 days per week, otherwise was occasional.

20 %: >19) groups. We found that their 10-year cumulative incidence of CRC was 0.24 %, 0.58 %, and 1.59 %, respectively (Fig. 2).

3.4. Performance of CRC risk prediction score in screening

In the Zhejiang screening programme, participants diagnosed with CRC had higher risk scores (mean ± SD: 20.9 ± 4.9) than those without CRC (18.1 ± 5.6) (Fig. 3A). The risk score strongly correlated with the detection rate of CRC, with an R^2 of 0.90 (Fig. 3B). The detection rates of CRC and advanced adenoma were 1.29 % (1185/91,575) and 11.3 % (10,313/91,575), respectively, both of which increased progressively as the score increased across ten groups by the decile cutoffs derived from the CKB (Fig. 3C and Supplemental Table 9). Specifically, the detection rates of CRC were 0.25 %, 0.82 %, and 1.93 % in low-risk (score <6), intermediate-risk (score: 6–19), and high-risk (score >19) groups, respectively. The corresponding rates of advanced adenoma were 3.15 %, 8.08 %, and 15.6 %, respectively (Supplemental Table 10).

The new score had a C-statistic of 0.65 (95 % CI: 0.63–0.66) for CRC, which was higher than that of the APCS score (0.60, 95 % CI: 0.58–0.61), the modified APCS score (0.60, 95 % CI: 0.59–0.62), and the KCS score (0.60, 95 % CI: 0.59–0.62) (Fig. 3D). Additionally, the new score had a higher NRI when compared to the APCS score ($NRI^{>0}$ = 36.9 %, $P < 0.0001$), the modified APCS score ($NRI^{>0}$ = 25.9 %, $P < 0.0001$), and the KCS score ($NRI^{>0}$ = 32.4 %, $P < 0.0001$) (Supplemental Table 11).

3.5. An online CRC risk prediction tool

We presented the new scoring method online as an easily and freely available tool named CORES (http://ccra.njmu.edu.cn/cores/web), which can be utilized by the general population to estimate their risk of CRC. Personalized recommendations for lifestyle changes and screening options are also provided according to individual's risk level.

4. Discussion

Leveraging data from the largest perspective cohort in China, we developed a CRC prediction risk score incorporating age, sex, BMI, smoking, alcohol consumption, red meat intake, prevalent diabetes, history of gallstone disease, and family history of cancer in first-degree relatives. Further evaluation in a large community-based screening programme demonstrated that the novel risk score outperformed commonly used models in terms of risk discrimination for CRC. These findings suggest that the risk score has the potential to identify individuals

Table 2
Selected predictors for colorectal cancer risk and corresponding risk points in the CKB cohort.

Variable	Cases/Person-years	Regression coefficient	HR (95 % CI)	*P* value	Points assigned
Age, years					
40-<45	235/912,354		1 (reference)		0
45-<50	298/742,413	0.39	1.48 (1.25–1.76)	<0.0001	5
50-<55	523/927,872	0.73	2.08 (1.78–2.42)	<0.0001	9
55-<60	561/716,881	1.06	2.88 (2.47–3.36)	<0.0001	12
60-<65	532/486,870	1.38	3.98 (3.41–4.65)	<0.0001	16
65-<70	548/380,022	1.64	5.14 (4.40–6.00)	<0.0001	19
70–75	439/247,420	1.85	6.34 (5.40–7.44)	<0.0001	22
Sex					
Female	1590/2,640,173		1 (reference)		0
Male	1546/1,773,660	0.21	1.24 (1.14–1.34)	<0.0001	2
BMI, kg/m^2					
<28	2681/3,922,234		1 (reference)		0
≥28	455/491,599	0.24	1.27 (1.15–1.41)	<0.0001	3
Pack-years of smoking					
<30	2559/3,902,897		1 (reference)		0
≥30	577/510,936	0.22	1.24 (1.12–1.38)	<0.0001	3
Alcohol drinking					
Occasional	2604/3,876,804		1 (reference)		0
Frequent	532/537,029	0.21	1.23 (1.11–1.37)	<0.0001	2
Red meat intake					
Occasional	1465/2,326,673		1 (reference)		0
Frequent	1671/2,087,160	0.12	1.13 (1.05–1.22)	0.002	1
Prevalent diabetes					
No	2820/4,138,808		1 (reference)		0
Yes	316/275,025	0.20	1.23 (1.09–1.38)	0.001	2
History of gallstone disease					
No	2881/4,127,223		1 (reference)		0
Yes	255/286,610	0.14	1.15 (1.01–1.31)	0.04	2
Family history of cancer in first-degree relatives					
No	2490/3,587,917		1 (reference)		0
Yes	646/825,916	0.09	1.09 (1.00–1.19)	0.05	1

Abbreviations: CKB, China Kadoorie Biobank; HR, hazard ratio; CI, confidence interval; BMI, body mass index.

at high risk of CRC, which could improve screening efficiency and facilitate tailored prevention efforts.

To date, there have been only a few risk prediction models developed in the Chinese population. In a cross-sectional study involving 7541 individuals who underwent routine colonoscopy at 19 hospitals, Cai et al. established a model for advanced colorectal neoplasia based on eight variables [8]. The variables included age, sex, smoking, diabetes, green vegetables, pickled food, fried food, and white meat, achieving a C-statistic of 0.74. However, the lack of prospective evaluation and external validation of the model raises concerns about recall bias in the ascertainment of risk factors and selection bias due to the hospital-based population [8]. To the best of our knowledge, only one prior study created a CRC risk model based on a prospective cohort in China, which included 92,923 men from the Kailuan Cohort Study with a 10-year follow-up and 353 incident CRC cases [11]. The model incorporated age, alcohol consumption, waist circumference, occupational sitting time, and diabetes, showing a C-statistic of 0.66 [11]. However, the study was limited the lack of external validation, absence of information on family history of CRC, and inclusion of only men. In the current study, we created a new risk score based on a nationwide prospective cohort and evaluated the score's performance in an independent screening programme. The large sample size and rigorous design ensured an improved performance of the new score in estimating an individual's CRC risk. Consistent with previous evidence [30], our multivariable analyses showed that age, sex, BMI, smoking, alcohol consumption, red meat intake, prevalent diabetes, and family history of cancer were independent risk factors of CRC. Moreover, we identified a positive association between history of gallstone disease and CRC risk, which aligns with the results from previous observational and Mendelian randomization analyses [31,32]. Although the exact mechanism is unknown, it has been suggested that altered bile acid metabolism and gut microbiota dysbiosis may be underlying factors and thus warrant further investigation [33].

The APCS score was developed in 2011 as a tool to identify individuals at high risk of advanced colorectal neoplasia in asymptomatic Asian populations [26]. Through a cross-sectional study that included a derivation set of 860 subjects and a validation set of 1892 subjects from 11 Asian cities, the score was established by incorporating age, sex, family history of CRC, and smoking status [26]. Other risk factors, such as obesity, alcohol consumption, red meat intake, and diabetes were overlooked, despite their association with increased risk of CRC in previous studies [34]. Several subsequent studies in Asia attempted to improve the predictive ability of the APCS score by incorporating additional predictors. For instance, both the modified APCS score [27] and the KCS score [28] added BMI with different scoring methods and yielded better discrimination. In the current study, we evaluated the potential utility of a new risk score that incorporated additional predictors such as alcohol consumption, red meat intake, diabetes, and history of gallstone disease. We found that the new score exhibited improved discrimination for CRC compared to previous scores, supporting the inclusion of these additional risk factors in the Chinese population.

In addition to existing risk prediction scores, several novel screening techniques such as multi-target stool DNA tests, gut microbiota assays, and blood-based methylated DNA tests show promise in optimizing risk-adapted screening [6]. Incorporating genetic susceptibility biomarkers has also demonstrated improved predictive power [35]. These approaches have the potential to attract individuals who are currently hesitant to participate in colonoscopy screening. However, the new techniques have high dependence on specialized equipment and technicians, leading to additional costs and complexity for large-scale screening. In contrast, the new risk score is easy measurable and available online, making it a convenient tool for risk-adapted screening and promoting healthy lifestyles to reduce CRC incidence. Nonetheless, there is always room for improvement in risk assessment accuracy. Further studies are necessary to determine if incorporating potential biomarkers in the risk score can enhance its performance, and to evaluate its feasibility and cost-effectiveness in practice.

The current study has several strengths. First, we developed a new risk prediction score within the largest nationwide prospective cohort in China, ensuring the robustness and potential generalizability of the score to the Chinese population. Second, we evaluated the score in a large-scale screening programme, supporting its use as an adjunct to a risk-adapted screening in China. However, there are also potential limitations to consider. Self-reported information on lifestyle, diet, and medical history may be subject to measurement errors, although a quality control survey with 15,728 participants in the CKB indicated good

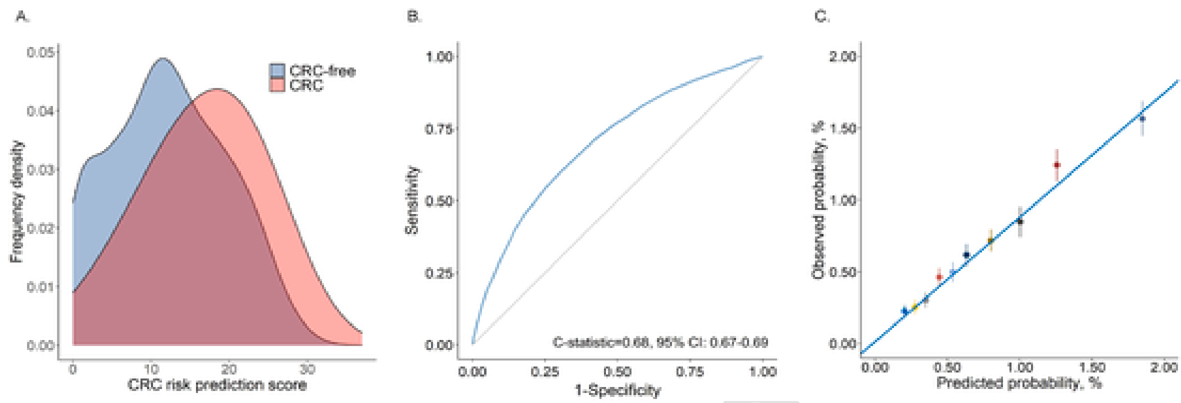

Fig. 1. Distribution and calibration of colorectal cancer risk prediction score in the China Kadoorie Biobank cohort. A. Distribution of the score between incident CRC cases and CRC-free participants. B. Receiver operating characteristic curve at 10 years. C. Calibration plot of the predicted and observed 10-year probability of CRC by the score deciles.

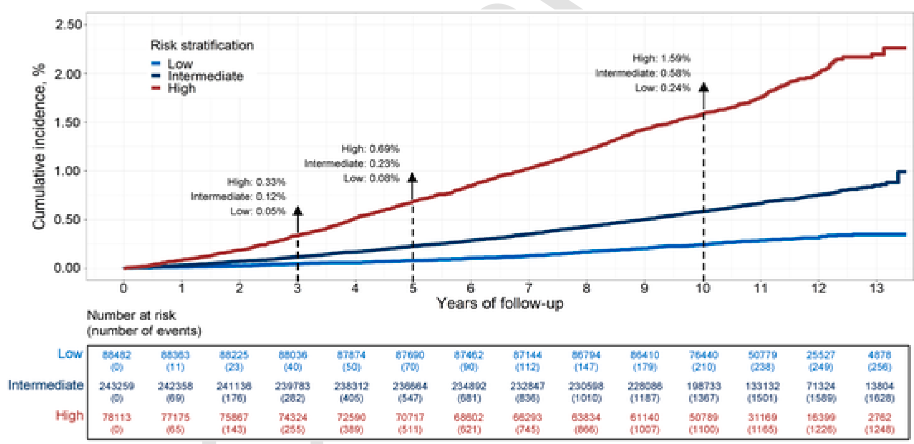

Fig. 2. Cumulative incidence plot of colorectal cancer in the China Kadoorie Biobank cohort. Participants were divided into low-risk (bottom 20 % of the score: <6), intermediate-risk (20%–80 %: 6–19), and high-risk (top 20 %: >19) groups. Cumulative incidence of CRC was calculated by using the Kaplan-Meier method. The risk table under plot showed the number at risk and the corresponding cumulative number of incident CRC cases at years of follow-up.

reproducibility for common variables [13]. Additionally, the CKB did not collect data on family history of CRC, and we used family history of all cancers as a surrogate to capture the inherited susceptibility of CRC as much as possible.

In conclusion, our two-stage multicenter design with a large number of participants allowed us to establish a novel risk prediction score that is useful to identify individuals at differential risk of CRC and triage high-risk individuals for screening. The web-based tool is easily implemented and freely accessible to the general population, providing a cornerstone for a tailored screening service that promotes early prevention of CRC in China.

Ethics approval and consent to participate

All participants signed a written informed consent on enrollment. Ethics approvals of the China Kadoorie Biobank were obtained from Oxford University (025–04) and the Chinese Center for Disease Control and Prevention (005/2004). The Zhejiang CRC screening programme was approved by the Ethics Committee and Institutional Review Board of Zhejiang Cancer Hospital (ID: IRB-2023-464).

Funding

The CKB baseline survey and the first re-survey were supported by a grant from the Kadoorie Charitable Foundation in Hong Kong. The long-term follow-up is supported by grants from the National Natural Science Foundation of China (82192900, 81390540, 91846303, 81941018), grants (2016YFC0900500) from the National Key R&D Program of China, Chinese Ministry of Science and Technology (2011BAI09B01), and UK Wellcome Trust (212946/Z/18/Z, 202922/Z/16/Z, 104085/Z/14/Z, 088158/Z/09/Z). This research was also supported by the National Natural Science Foundation of China (81973127, 81922061, 81521004), CAMS Innovation Fund for Medical Sciences (2019RU038), Special Foundation for National Science and Technology Basic Research Program of China (2019FY101104), and Outstanding Youth Fund of Jiangsu Natural Science Foundation (BK20230005). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

Details of how to access the China Kadoorie Biobank data are available from https://www.ckbiobank.org/data-access. The application number of this research is DAR-2022-00372. Deidentified individual

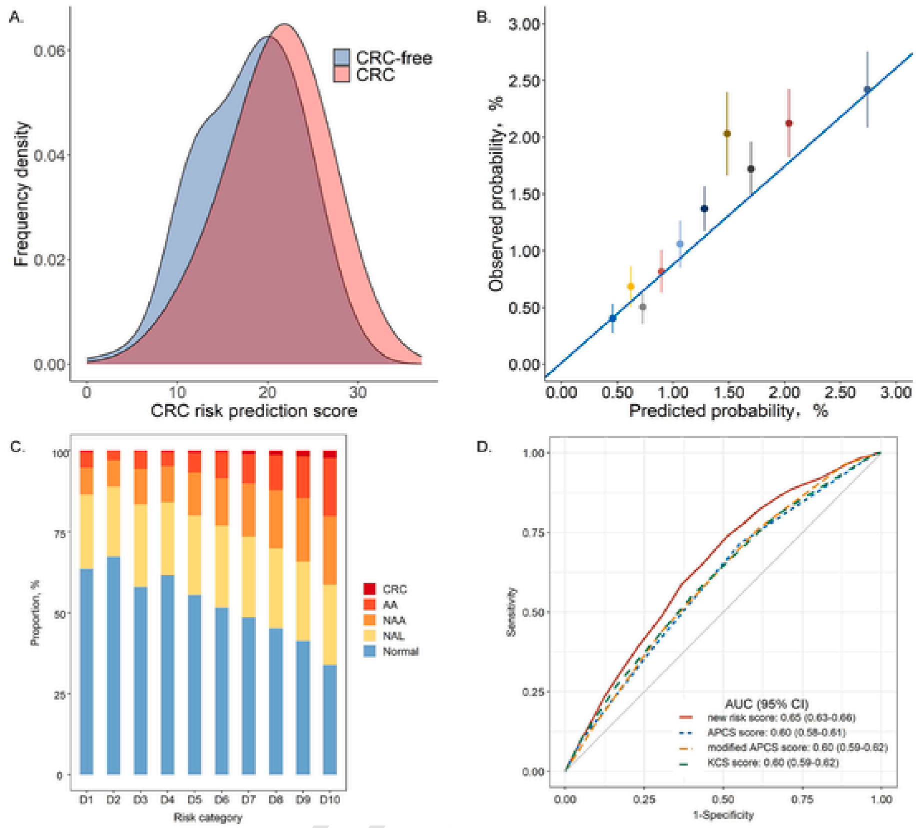

Fig. 3. Performance of the newly developed score in the Zhejiang CRC screening programme. A. Distribution of the score between CRC cases and CRC-free participants; B. Calibration plot of the CRC detection rate; C. Proportions of different colorectal lesions in each risk category. Ten risk categories (D1-D10) were based on the same cut-offs of the score deciles from the China Kadoorie Biobank cohort; D. Receiver operating characteristic curves for CRC based on the new score, the APCS score, the modified APCS score, and the KCS score. Abbreviations: AA, advanced adenoma; APCS, Asia-Pacific colorectal screening; AUC, area under the curve; CRC, colorectal cancer; KCS, Korean colorectal screening; NAA, non-advanced adenoma; NAL, non-adenomatous lesion.

participant datasets of the Zhejiang screening programme are available from the corresponding author upon reasonable request.

CRediT authorship contribution statement

Dong Hang: Writing – original draft, Methodology, Formal analysis, Conceptualization. **Dianjianyi Sun:** Validation, Data curation. **Lingbin Du:** Investigation, Data curation. **Jianv Huang:** Data curation. **Jiacong Li:** Investigation. **Chen Zhu:** Investigation. **Le Wang:** Investigation. **Jingjing He:** Investigation. **Xia Zhu:** Writing – review & editing. **Meng Zhu:** Writing – review & editing. **Ci Song:** Writing – review & editing. **Juncheng Dai:** Writing – review & editing. **Canqing Yu:** Writing – review & editing, Project administration. **Zekuan Xu:** Resources. **Ni Li:** Writing – review & editing. **Hongxia Ma:** Writing – review & editing, Resources. **Guangfu Jin:** Writing – review & editing, Resources. **Ling Yang:** Data curation. **Yiping Chen:** Data curation. **Huaidong Du:** Data curation. **Xiangdong Cheng:** Writing – review & editing. **Zhengming Chen:** Resources, Funding acquisition. **Jun Lv:** Writing – review & editing, Resources, Funding acquisition. **Zhibin Hu:** Writing – review & editing, Conceptualization. **Liming Li:** Writing – review & editing, Supervision, Funding acquisition. **Hongbing Shen:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The chief acknowledgement is to the participants in the CKB and the members of the survey teams in each of the 10 regional centres, as well as to the project development and management teams based at Beijing, Oxford and the 10 regional centres. We also thank the study participants and research staff for their contributions in the Zhejiang CRC screening programme.

Abbreviations

CRC	Colorectal cancer
CKB	China Kadoorie Biobank
BMI	Body mass index
FIT	Fecal immunochemical testing

AA	Advanced adenoma
HRs	Hazards ratios
CIs	Confidence intervals
AUC	Area under the receiver operating characteristic curve
APCS	Asia-Pacific Colorectal Screening
KCS	Korean Colorectal Screening
NRI	Net reclassification improvement
NAA	Non-advanced adenoma
NAL	Non-adenomatous lesion

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.canlet.2024.217057.

References

[1] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA A Cancer J. Clin. 71 (2021) 209–249.

[2] B. Lu, N. Li, C.Y. Luo, J. Cai, M. Lu, Y.H. Zhang, H.D. Chen, M. Dai, Colorectal cancer incidence and mortality: the current status, temporal trends and their attributable risk factors in 60 countries in 2000-2019, Chin. Med. J. 134 (2021) 1941–1951.

[3] Y. Yang, Z. Han, X. Li, A. Huang, J. Shi, J. Gu, Epidemiology and risk factors of colorectal cancer in China, Chinese journal of cancer research = Chung-kuo yen cheng yen chiu 32 (2020) 729–741.

[4] H. Chen, N. Li, J. Ren, X. Feng, Z. Lyu, L. Wei, X. Li, L. Guo, Z. Zheng, S. Zou, Y. Zhang, J. Li, K. Zhang, W. Chen, M. Dai, J. He, C. group of Cancer Screening Program in Urban, Participation and yield of a population-based colorectal cancer screening programme in China, Gut 68 (2019) 1450–1457.

[5] C. Xia, X. Dong, H. Li, M. Cao, D. Sun, S. He, F. Yang, X. Yan, S. Zhang, N. Li, W. Chen, Cancer statistics in China and United States, 2022: profiles, trends, and determinants, Chin. Med. J. 135 (2022) 584–590.

[6] M.A. Hull, C.J. Rees, L. Sharp, S. Koo, A risk-stratified approach to colorectal cancer prevention and diagnosis, Nature reviews, Gastroenterol. Hepatol. 17 (2020) 773–780.

[7] J.A. Usher-Smith, F.M. Walter, J.D. Emery, A.K. Win, S.J. Griffin, Risk prediction models for colorectal cancer: a systematic review, Cancer Prev. Res. 9 (2016) 13–26.

[8] Q.C. Cai, E.D. Yu, Y. Xiao, W.Y. Bai, X. Chen, L.P. He, Y.X. Yang, P.H. Zhou, X.L. Jiang, H.M. Xu, H. Fan, Z.Z. Ge, N.H. Lv, Z.G. Huang, Y.M. Li, S.R. Ma, J. Chen, Y.Q. Li, J.M. Xu, P. Xiang, L. Yang, F.L. Lin, Z.S. Li, Derivation and validation of a prediction rule for estimating advanced colorectal neoplasm risk in average-risk Chinese, Am. J. Epidemiol. 175 (2012) 584–593.

[9] G. Chen, B. Mao, Q. Pan, Q. Liu, X. Xu, Y. Ning, Prediction rule for estimating advanced colorectal neoplasm risk in average-risk populations in southern Jiangsu Province, Chinese journal of cancer research = Chung-kuo yen cheng yen chiu 26 (2014) 4–11.

[10] J. Shen, Y. Wu, X. Feng, F. Liang, M. Mo, B. Cai, C. Zhou, Z. Wang, M. Zhu, G. Cai, Y. Zheng, Assessing individual risk for high-risk early colorectal neoplasm for pre-selection of screening in Shanghai, China: a population-based nested case-control study, Cancer Manag. Res. 13 (2021) 3867–3878.

[11] L. Guo, H. Chen, G. Wang, Z. Lyu, X. Feng, L. Wei, X. Li, Y. Wen, M. Lu, Y. Chen, J. Shi, J. Ren, C. Lin, X. Yu, S. Chen, S. Wu, N. Li, M. Dai, J. He, Development of a risk score for colorectal cancer in Chinese males: a prospective cohort study, Cancer Med. 9 (2020) 816–823.

[12] Z. Chen, L. Lee, J. Chen, R. Collins, F. Wu, Y. Guo, P. Linksted, R. Peto, Cohort profile: the Kadoorie study of chronic disease in China (KSCDC), Int. J. Epidemiol. 34 (2005) 1243–1249.

[13] Z. Chen, J. Chen, R. Collins, Y. Guo, R. Peto, F. Wu, L. Li, g. China Kadoorie Biobank collaborative, China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up, Int. J. Epidemiol. 40 (2011) 1652–1666.

[14] P.C. Austin, E.A. Stuart, Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies, Stat. Med. 34 (2015) 3661–3679.

[15] N. Keum, E. Giovannucci, Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies, Nat. Rev. Gastroenterol. Hepatol. 16 (2019) 713–732.

[16] T. Sawicki, M. Ruszkowska, A. Danielewicz, E. Niedźwiedzka, T. Arłukowicz, K.E. Przybyłowicz, A review of colorectal cancer in terms of epidemiology, risk factors, development, symptoms and diagnosis, Cancers 13 (2021).

[17] M.C. Tammemagi, H.A. Katki, W.G. Hocking, T.R. Church, N. Caporaso, P.A. Kvale, A.K. Chaturvedi, G.A. Silvestri, T.L. Riley, J. Commins, C.D. Berg, Selection criteria for lung-cancer screening, N. Engl. J. Med. 368 (2013) 728–736.

[18] Y.R. Li, J. Wang, L.Y. Zhao, Z.H. Wang, D.M. Yu, Y.N. He, G.G. Ding, [The drinking status and associated factors in adults in China], Zhonghua liu xing bing xue za zhi = Zhonghua liuxingbingxue zazhi 39 (2018) 898–903.

[19] H. Du, Y. Guo, D.A. Bennett, F. Bragg, Z. Bian, M. Chadni, C. Yu, Y. Chen, Y. Tan, I.Y. Millwood, W. Gan, L. Yang, P. Yao, G. Luo, J. Li, Y. Qin, J. Lv, X. Lin, T. Key, J. Chen, R. Clarke, L. Li, Z. Chen, Red meat, poultry and fish consumption and risk of diabetes: a 9 year prospective cohort study of the China Kadoorie Biobank, Diabetologia 63 (2020) 767–779.

[20] B.F. Zhou, C. Cooperative Meta-Analysis Group of the Working Group on Obesity in, Predictive values of body mass index and waist circumference for risk factors of certain related diseases in Chinese adults--study on optimal cut-off points of body mass index and waist circumference in Chinese adults, Biomed. Environ. Sci. 15 (2002) 83–96.

[21] D.A. Lieberman, D.K. Rex, S.J. Winawer, F.M. Giardiello, D.A. Johnson, T.R. Levin, Guidelines for colonoscopy surveillance after screening and polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer, Gastroenterology 143 (2012) 844–857.

[22] B. Click, P.F. Pinsky, T. Hickey, M. Doroudi, R.E. Schoen, Association of colonoscopy adenoma findings with long-term colorectal cancer incidence, JAMA 319 (2018) 2021–2031.

[23] D. Schoenfeld, Partial residuals for the proportional hazards regression model, Biometrika 69 (1982) 239–241.

[24] H.B. Mehta, V. Mehta, C.J. Girman, D. Adhikari, M.L. Johnson, Regression coefficient-based scoring system should be used to assign weights to the risk index, J. Clin. Epidemiol. 79 (2016) 22–28.

[25] D.C. Muller, M. Johansson, P. Brennan, Lung cancer risk prediction model incorporating lung function: development and validation in the UK Biobank prospective cohort study, J. Clin. Oncol. 35 (2017) 861–869.

[26] K.G. Yeoh, K.Y. Ho, H.M. Chiu, F. Zhu, J.Y. Ching, D.C. Wu, T. Matsuda, J.S. Byeon, S.K. Lee, K.L. Goh, J. Sollano, R. Rerknimitr, R. Leong, K. Tsoi, J.T. Lin, J.J. Sung, C. Asia-Pacific Working Group on Colorectal, The Asia-Pacific Colorectal Screening score: a validated tool that stratifies risk for colorectal advanced neoplasia in asymptomatic Asian subjects, Gut 60 (2011) 1236–1241.

[27] J.J.Y. Sung, M.C.S. Wong, T.Y.T. Lam, K.K.F. Tsoi, V.C.W. Chan, W. Cheung, J.Y.L. Ching, A modified colorectal screening score for prediction of advanced neoplasia: a prospective study of 5744 subjects, J. Gastroenterol. Hepatol. 33 (2018) 187–194.

[28] D.H. Kim, J.M. Cha, H.P. Shin, K.R. Joo, J.I. Lee, D.I. Park, Development and validation of a risk stratification-based screening model for predicting colorectal advanced neoplasia in Korea, J. Clin. Gastroenterol. 49 (2015) 41–49.

[29] M.J. Pencina, R.B. D'Agostino Sr, R.B. D'Agostino Jr, R.S. Vasan, Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond, Stat. Med. 27 (2008) 157–172 ; discussion 207-112.

[30] F. Islami, A. Goding Sauer, K.D. Miller, R.L. Siegel, S.A. Fedewa, E.J. Jacobs, M.L. McCullough, A.V. Patel, J. Ma, I. Soerjomataram, W.D. Flanders, O.W. Brawley, S.M. Gapstur, A. Jemal, Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States, CA A Cancer J. Clin. 68 (2018) 31–54.

[31] D.M. Shabanzadeh, L.T. Sorensen, T. Jorgensen, Association between screen-detected gallstone disease and cancer in a cohort study, Gastroenterology 152 (2017) 1965–1974 e1961.

[32] Y. Pang, J. Lv, C. Kartsonaki, Y. Guo, C. Yu, Y. Chen, L. Yang, Z. Bian, I.Y. Millwood, R.G. Walters, X. Li, J. Zou, M.V. Holmes, J. Chen, Z. Chen, L. Li, Causal effects of gallstone disease on risk of gastrointestinal cancer in Chinese, Br. J. Cancer 124 (2021) 1864–1872.

[33] K. Soreide, Gallstone disease and cancer risk: finding the bug in the system, Gastroenterology 152 (2017) 1825–1828.

[34] F. Kastrinos, S.S. Kupfer, S. Gupta, Colorectal cancer risk assessment and precision approaches to screening: brave new world or worlds apart? Gastroenterology 164 (2023) 812–827.

[35] L. McGeoch, C.L. Saunders, S.J. Griffin, J.D. Emery, F.M. Walter, D.J. Thompson, A.C. Antoniou, J.A. Usher-Smith, Risk prediction models for colorectal cancer incorporating common genetic variants: a systematic review, Cancer Epidemiol. Biomarkers Prev. 28 (2019) 1580–1593.