Medicine®

OPEN

Predicting the risk for colorectal cancer with personal characteristics and fecal immunochemical test

Wen Li, MD[a,b], Li-Zhong Zhao, MD[a,b], Dong-Wang Ma, MD[a,b], De-Zheng Wang, MD[c], Lei Shi, MD[a,b], Hong-Lei Wang, MD[a,b], Mo Dong, MD[a,b], Shu-Yi Zhang, MD[a,b], Lei Cao, MD[a,b], Wei-Hua Zhang, MD[a,b], Xi-Peng Zhang, MD[a,b], Qing-Huai Zhang, MD[a,b], Lin Yu, MD[a,b], Hai Qin, MD[a,b], Xi-Mo Wang, MD[a,d,*], Sam Li-Sheng Chen, PhD[e,*]

Abstract

We aimed to predict colorectal cancer (CRC) based on the demographic features and clinical correlates of personal symptoms and signs from Tianjin community-based CRC screening data.

A total of 891,199 residents who were aged 60 to 74 and were screened in 2012 were enrolled. The Lasso logistic regression model was used to identify the predictors for CRC. Predictive validity was assessed by the receiver operating characteristic (ROC) curve. Bootstrapping method was also performed to validate this prediction model.

CRC was best predicted by a model that included age, sex, education level, occupations, diarrhea, constipation, colon mucosa and bleeding, gallbladder disease, a stressful life event, family history of CRC, and a positive fecal immunochemical test (FIT). The area under curve (AUC) for the questionnaire with a FIT was 84% (95% CI: 82%–86%), followed by 76% (95% CI: 74%–79%) for a FIT alone, and 73% (95% CI: 71%–76%) for the questionnaire alone. With 500 bootstrap replications, the estimated optimism (<0.005) shows good discrimination in validation of prediction model.

A risk prediction model for CRC based on a series of symptoms and signs related to enteric diseases in combination with a FIT was developed from first round of screening. The results of the current study are useful for increasing the awareness of high-risk subjects and for individual-risk-guided invitations or strategies to achieve mass screening for CRC.

Abbreviations: AUC = area under curve, CRC = colorectal cancer, FIT = fecal immunochemical test, QFIT = questionnaire and fecal immunochemical test, ROC = receiver operating characteristic.

Keywords: colorectal cancer, fecal immunochemical test, prediction, screening

1. Introduction

There are numerous risk prediction models for colorectal cancer (CRC) in the literature,[1,2] including both genetic[3–10] and non-genetic models.[11–22] The former only ascertains information

[a] Department of Epidemiology, Tianjin Colorectal and Anal Disease Research Institute, [b] Department of Gastroenterology, Tianjin Union Medical Center, [c] Non-Communicable Disease Control and Prevention, Tianjin Centers for Disease Control and Prevention, [d] Department of Gastroenterology, Tianjin Nankai Hospital, Tianjin, P.R. China, [e] School of Oral Hygiene, College of Oral Medicine, Taipei Medical University, Taiwan.

* Correspondence: Xi-Mo Wang, Department of Gastroenterology, Tianjin Nankai Hospital, No.6 Changjiang Road Nankai District, Tianjin 300100, P.R. China (e-mail: ximowang12@163.com); Sam Li-Sheng Chen, School of Oral Hygiene, College of Oral Medicine, Taipei Medical University, No.250, Wuxing St., Taipei 11031, Taiwan (e-mail: samchen@tmu.edu.tw).

from a small proportion of population that has inherited a genetic mutation related to several major genes, such as the MMR genes, APC, MUTYH, STK11 (LB1), BMPR1A, and PTEN. The most notable model is the MMRpro model, which considers the dominantly inherited, highly penetrant mutations (MLH1, MSH2, or MSH6).[3] The non-genetic models aim to identify high-risk subjects based on non-genetic, personal, and environmental factors together with a family history of CRC. Although these models are probably used to identify subjects with a high risk of CRC susceptibility, there are still some concerns over their usefulness. Genetic models can only identify half of the familial risk of CRC,[23] and the predictive accuracy of non-genetic models is modest.[20–21]

To increase the predictive validity of these prediction models, a history of screening or diagnosis with colonoscopy has been incorporated into the risk prediction model, according to previous studies.[14–16] In addition to the well-established risk factors, adding information related to the awareness variables might help capture the risk of CRC susceptibility. This aspect can partially account for an ethnic-specific incidence of CRC worldwide because the extent of awareness may affect the detection and early treatment of CRC, as can genetic differences. Adding this information also plays an important role in individual-risk-guided community-based screening. CRC awareness is still low in Asian countries compared with Western countries. Self-reported symptoms and signs related to enteric disease may provide useful information regarding the economic influence of early CRC detection in a country with low

awareness. Although these clinical symptoms and signs may not be associated with CRC, they may create an incentive for being wary of seeking medical care to detect CRC earlier.

In addition to considering the history of screening by colonoscopy, the use of a fecal immunochemical test (FIT) may provide valuable information,[24,25] as a previous study suggested performing a colonoscopy to detect colorectal cancer. How to incorporate information on the history of FIT into a risk prediction model should also be considered to improve awareness.

Data from population-based screening for CRC with a questionnaire and fecal immunochemical test (QFIT), in Tianjin, offer an opportunity to test whether and how information on the personal characteristics of the symptoms and signs related to enteric diseases obtained from the questionnaire together with FIT are associated with the risk for CRC. The aims of this study were to examine the association between a constellation of demographic features and clinical correlates on the personal symptoms and signs related to enteric diseases. Those correlates that were identified as significant were used to compute the individual risk score and build a risk prediction model for colorectal cancer.

2. Methods

2.1. Study samples

We recruited 891,199 community residents aged 60 to 74 years who participated in a Tianjin community-based CRC screening program in China in 2012. Information on CRC cases consisted of 207 screen-detected CRC cases and 264 clinically detected CRC cases identified from the Tianjin Union Medical Center during 1 year of follow-up of the entire screened cohort.

2.2. Screening protocol and mass screening

The procedure of screening began with a questionnaire that contained personal characteristics and a constellation of clinical correlates. Those subjects who were defined as high-risk based on the questionnaire (53,631 out of 891,199) were referred to arrange further colonoscopic examinations. Those who were defined as low-risk based on the questionnaire were suggested to undergo a FIT (ABON, China). Due to the evaluation purpose of the pilot phase, a proportion of subjects returning a high-risk questionnaire also had a FIT performed to provide information for evaluating the sequential design. Since FIT is expected to be more accurate in detecting CRC than the questionnaire is, we had approximately one-third of the subjects identified as high-risk based on the questionnaire (20,633 out of 53,631) undergo a FIT and approximately two-thirds of the low-risk subjects based on the questionnaire (533,449 out of 891,199) undergo a FIT. Note that only a fraction of positive subjects was referred for colonoscopy.

2.3. Questionnaire

A structured questionnaire that included personal characteristics and a constellation of clinical correlates was provided during a screening activity. We examined the demographic information, including age, sex, marital status, education level, and occupation. We also measured the clinical correlates based on the following nine questions: chronic diarrhea history; chronic constipation history; mucus or blood stool history; chronic appendicitis or appendectomy; chronic gallbladder disease or

gallbladder surgery history; stressful life event over the past 2 decades; cancer history; colon polyps history; or family history of colorectal cancer among first-degree relatives. The questionnaire was performed by trained public health nurses.

Subjects who had any first-degree relatives with CRC cancer, who had ever been affected by polyps or cancer or who had ≥2 of the following clinical syndromes, that is, chronic constipation, chronic diarrhea, bloody mucus, history of negative life events, history of chronic appendicitis or appendectomy, history of chronic gallbladder disease or gallbladder surgery, were defined as high-risk subjects. Those who were defined as high-risk subjects based on the questionnaire were referred for further colonoscopic examination. Those who were defined as low-risk subjects based on the questionnaire were suggested to undergo a FIT. Subjects with positive findings on the FIT were also referred for colonoscopic examination.

2.4. Immunochemical test

Fecal samples were obtained from 550,318 subjects at their home using the collection kit provided by the manufacturer (ABON, China). Participants were asked to collect 10 to 50 mg of stool and send it to the community hospitals. No specific dietary restriction was stipulated. All tests were processed at the laboratory within 8 hours after collection. According to the manufacturer's instructions, this qualitative test is considered positive when the sample is positive for hemoglobin. The results were reported by the central laboratory in a qualitative manner (positive and negative). Finally, 4% of the stool samples were randomly selected for a re-test quality control.

2.5. Ethical consideration

The original research protocol was reviewed and approved by the ethical review committee of Tianjin Union Medical Center. The program performs annual recruitment screenings, which are approved by the local ethical committee in the Health Bureau of Tianjin City. These approvals include data linkage systems and strict maintenance of participant confidentiality. Because the personal identification numbers for the datasets were encoded, the privacy and confidentiality of patients were ensured by obscuring the links between datasets.

2.6. Statistical Analysis

A Lasso logistic regression model was used to select potential predictors of CRC and to estimate the regression coefficients for the relationships between the predictors and CRC. The Hosmer–Lemeshow goodness-of-fit test was used to determine whether the prediction model was correctly specified. The calibration plots have the predicted probabilities for groups defined by ranges of individual predicted probabilities (10 groups of equal size) on the x-axis, and the mean observed outcome on the y-axis. These plots are graphical illustrations of the Hosmer–Lemeshow goodness-of-fit test. The receiver operating characteristic (ROC) curve analysis was applied for prediction. We determined the predictive ability for the prediction model by examining the area under the curve (AUC) using a non-parametric method such as the Mann–Whitney U test. We examined the models with questionnaire-based or FIT predictors. We also examined the model with both the questionnaire and FIT. The bootstrap method was adopted to validate the prediction model. We first developed our prediction models with the total sample and then generate a bootstrap sample by sampling n individuals with

replacement from the original sample. The sample size varied according to the number of events per variable (EPV). In order to develop an adequate predictive model, it has been suggested that EPV should be at least 10.[26] Thus, we present detailed results for EPV values starting from 5 to 80. The apparent performance was determined on random samples from the data set for EPV 5, 10, 20, 40, and 80, respectively. Simulations were repeated 500 times.

3. Results

3.1. Determination of predictors from model selections

The relevant factors included demographic factors (age, sex, education, and occupation); personal disease history of diarrhea, constipation, colon mucus and bleeding; appendicitis; occurrence of gallbladder disease; history of colon cancer and polyps; stressful life events; and family history. Table 1 shows the

Table 1

Distribution of characteristics for free-of-CRC versus CRC participants.

Characteristics	Category	Free of CRC		CRC		Total
		N	%	N	%	
Age (mean ± SD)		65.7 ± 4.2		66.7 ± 4.2		65.7 ± 4.2
Gender	Female	472,625	99.96	203	0.04	472,828
	Male	418,103	99.94	268	0.06	418,371
Marital status	Married	807,856	99.95	435	0.05	808,291
	Unmarried	9063	99.99	1	0.01	9064
	Divorced/widowed	73,396	99.96	31	0.04	73,427
	NK	413		4		417
Education	Illiterate	80,868	99.96	30	0.04	80,898
	Elementary School	330,708	99.97	102	0.03	330,810
	High School	410,941	99.93	274	0.07	411,215
	College or above	67,840	99.91	61	0.09	67,901
	NK	371		4		375
Occupation	Farmer	361,241	99.97	103	0.03	361,344
	Labor	293,288	99.94	186	0.06	293,474
	Personal Business	6063	99.98	1	0.02	6064
	Professional	100,570	99.91	89	0.09	100,659
	Other	129,153	99.93	88	0.07	129,241
	NK	413		4		417
Diarrhea	No	854,336	99.95	388	0.05	854,724
	Yes	36,293	99.78	80	0.22	36,373
	NK	99		3		102
Constipation	No	818,980	99.95	393	0.05	819,373
	Yes	71,650	99.90	75	0.10	71,725
	NK	98		3		101
Colon mucosa and bleeding	No	876,108	99.96	378	0.04	876,486
	Yes	14,520	99.38	90	0.62	14,610
	NK	100		3		103
Appendicitis	No	860,457	99.95	446	0.05	860,903
	Yes	30,169	99.93	22	0.07	30,191
	NK	102		3		105
Gallbladder disease	No	862,543	99.95	439	0.05	862,982
	Yes	28,085	99.90	29	0.10	28,114
	NK	100		3		103
Life stress event	No	821,142	99.95	428	0.05	821,570
	Yes	69,467	99.94	39	0.06	69,506
	NK	119		4		123
CRC history	No	884,962	99.95	458	0.05	885,420
	Yes	5556	99.84	9	0.16	5565
	NK	210		4		214
Colon polyps history	No	887,613	99.95	461	0.05	888,074
	Yes	2930	99.76	7	0.24	2937
	NK	185		3		188
CRC family history	No	883,387	99.95	446	0.05	883,833
	Yes	6809	99.68	22	0.32	6831
	NK	532		3		535
FIT	Negative	534,822	99.98	128	0.02	534,950
	Positive	15,185	98.81	183	1.19	15,368
	Not performed	340,721	99.95	160	0.05	340,881

CRC = colorectal cancer, NK = not known.

Table 2

Regression coefficients for predictors by model selection.

		Regression coefficients		
	Univariate		Multivariable	
Variables	Unadjusted model	Model 1 (questionnaire only)	Model 2 (FIT only)	Model 3 (Questionnaire + FIT)
N	891,199	89,0131	891,199	890,235
Number of event	471	465	471	466
Age	0.0529	0.0551	0.0493	0.0523
Gender (M vs F)	0.4003	0.3835	0.364	0.3231
Marital status				
Married versus Divorced/Widowed	0.6092	–	–	–
Unmarried versus Divorced/Widowed	−0.9756	–	–	–
Education				
Elementary school versus illiterate	−0.5064	−0.2669	–	−0.2657
High school versus illiterate	0.2645	0.1064	–	0.1115
College or above versus illiterate	0.5637	0.0806	–	0.0672
Occupation				
Labor versus farmer	0.7933	0.3309	–	0.3008
Personal business versus farmer	−0.5430	−0.9099	–	−0.744
Professionals versus farmer	1.1326	0.485	–	0.5017
Other occupation versus Farmer	0.8711	0.3771	–	0.4045
Personal disease history				
Diarrhea (Yes vs No)	1.5801	0.9935	–	0.7350
Constipation (Yes vs No)	0.7799	0.3878	–	0.1395
Colon mucosa and bleeding (Yes vs No)	2.6652	2.132	–	1.6493
Appendicitis (Yes vs No)	0.3414	–	–	–
Gallbladder disease (Yes vs No)	0.7082	0.1203	–	0.0768
Colon cancer history (Yes vs No)	1.1413	0.4618	–	–
Stress life event (Yes vs No)	−0.0743	0.1634	–	0.2101
CRC family history (Yes vs No)	1.8563	1.1206	–	0.9519
Colon polyps history (Yes vs No)	−1.5267	–	–	–
FIT positive versus FIT negative	3.9187	–	3.9024	3.6719
FIT (not performed) versus FIT negative	0.6737	–	0.6697	0.5334
Shrinkage		0.965	0.981	0.986
Model X^2		493.1	938.0	1268.0

CRC = colorectal cancer, FIT = fecal immunochemical test.

distribution of selected variables for non-CRC and CRC among the participants. The unadjusted association between each candidate predictor and the risk of colorectal cancer are given in Table 2. Table 2 also shows the estimated regression coefficients for the associations between each factor and the risk for colorectal cancer, after adjusting for confounding factors.

The selected predictors, including age, sex, education level, occupation, diarrhea, constipation, colon mucus and bleeding, gallbladder disease, personal colon cancer history, stressful life event, and a CRC family history, were included in model 1 (questionnaire only model). It is very interesting to note that those subjects who did not undergo a FIT were at a greater risk than were those who had undergone a FIT when the high-risk subjects, as defined by the questionnaire, were considered a separate risk group. Those subjects with a positive FIT result had an extremely high risk for CRC compared with those who had negative FIT results (model 2).

In addition to FIT, the predictors determined by model selection, including age, sex, education level, occupations, diarrhea, constipation, presence of colonic mucus and bleeding, gallbladder disease, stressful life event, and CRC family history, were included in model 3 (questionnaire plus FIT model).

Table 3 shows the estimated results for the variables relevant to the use of FIT, after adjusting for age and sex. The differences were all statistically significant ($P < .0001$).

3.2. Risk score for CRC

According to regression coefficients estimated from the logistic regression model (Questionnaire + FIT) as shown in Table 2. The individual risk score was built while considering FIT using the following equation:

Risk Score = 0.5 × Age + 3 × Gender (M:1;F:0) − 2.7 × Education (Elementary school:1;Illiteracy:0) + 1.1 × Education

Table 3

Comparison of areas under the curve (AUCs) for conventional risk predictors, FIT, or a combination.

Models	Mann–Whitney U test (95% CI)	Difference (95% CI)	P-value
Questionnaire only	0.732 (0.707–0.756)	Reference	
FIT only	0.764 (0.739–0.789)	0.0322 (−0.0006–0.0649)	.0540
Questionnaire + FIT	0.838 (0.817–0.860)	0.1067 (0.0834–0.1300)	<.0001

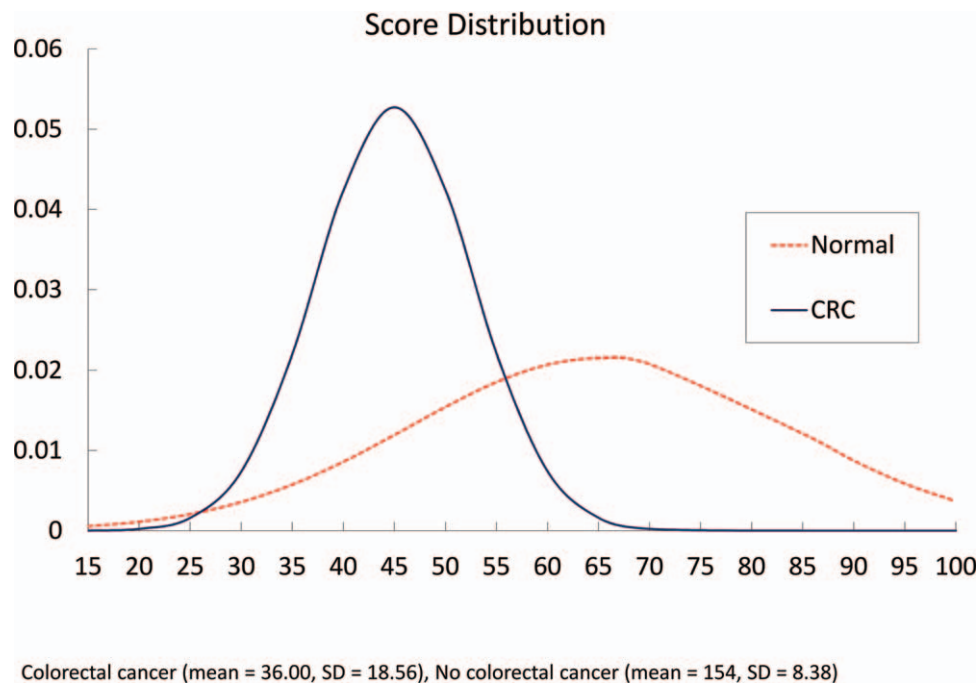CI = confidence interval, FIT = fecal immunochemical test.

Figure 1. Risk score distribution for Tianjin population.

(High school:1;Illiteracy:0) + 0.7 × Education (College or above:1; Illiteracy:0) + 3.0 × Labor (1 or 0) − 7.4 × Personal Business (1 or 0) + 5 × Professionals (1 or 0) + 4 × Other Occupation (1 or 0) + 7.4 × Diarrhea (1 or 0) + 14.0 × Constipation (1 or 0) + 16.5 × Colon Mucosa (1 or 0) + 7.7 × Gallbladder disease (1 or 0) + 2.1 × Stress life event(1 or 0) + 9.5 × CRC Family History (1 or 0) + 36.8 × FIT Positive (1 or 0) + 5.4 × FIT Not Perform (1 or 0).

The distribution of the risk scores between free-of-CRC and CRC subjects is shown in Fig. 1. The difference between the groups was statistically significant ($t = -23.38$, $P < .0001$).

3.3. Calibration

The comparison of the observed and predicted probabilities for the models according to the Hosmer–Lemeshow test is provided. The calibration plots are presented to reflect the agreement between observed outcomes and predictions (Fig. 2). The figures show perfect moderate calibration. The calibration lies on or around a 45° line of the plot. All of three developed prediction models show the good model calibration (Questionnaire only model: $\chi^2 = 11.77$; $P = .1617$, FIT only model: $\chi^2 = 9.38$; $P = .3116$, Questionnaire plus FIT model: $\chi^2 = 11.92$; $P = .1549$).

3.4. ROC analysis

The prediction models with significant correlates from the questionnaire plus FIT, FIT only, and the questionnaire only are also presented with ROC curves. Figure 3 shows that the questionnaire correlates in combination with FIT performed best among the 3 modes. The AUC (c-index) for the questionnaire with FIT was 0.838 (95% CI: 0.817–0.860), followed by 0.764 (95% CI: 0.739–0.789) for FIT only, and 0.732 (95% CI: 0.707–0.756) for the questionnaire only. A comparison between the full model and the model with FIT only, based on the non-parametric Mann–Whitney U test, is shown in Table 3.
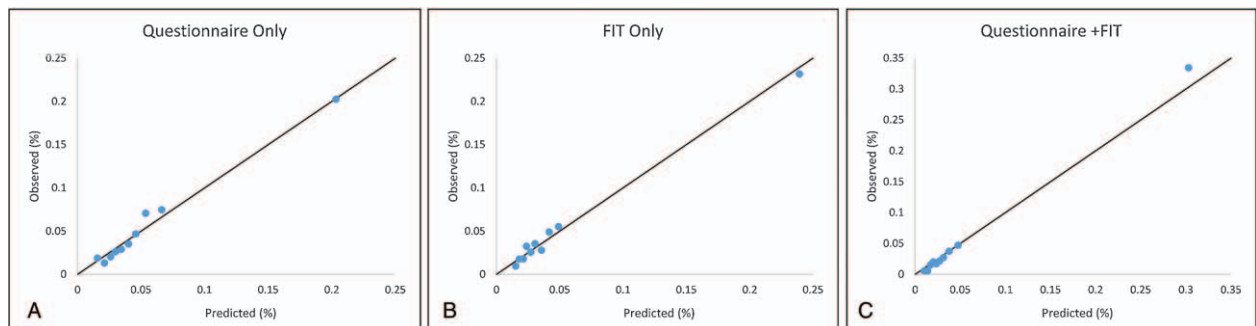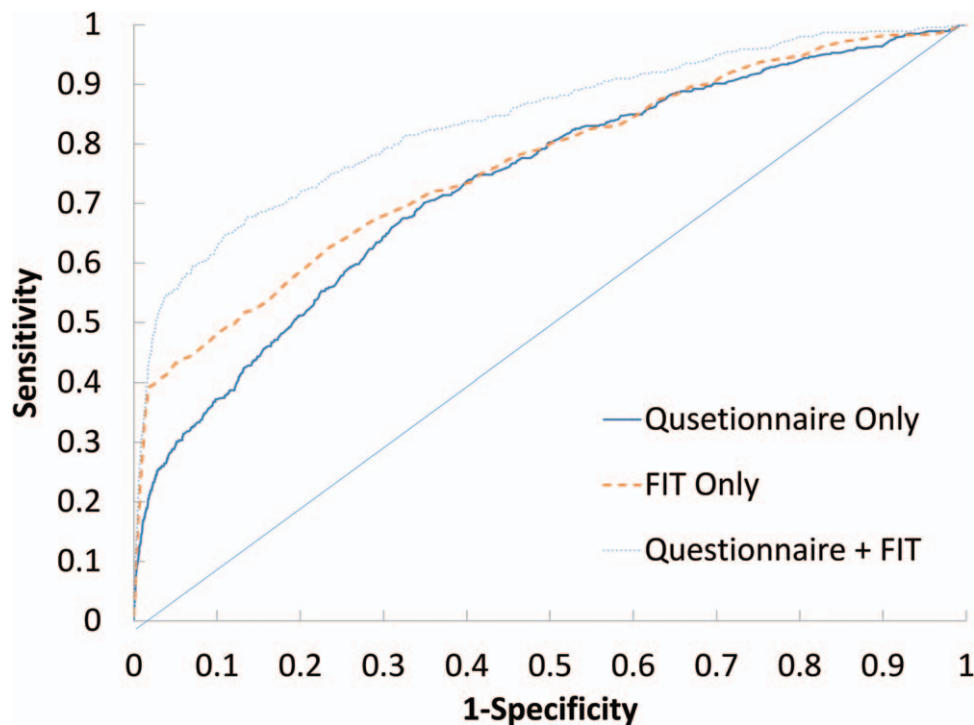

Figure 2. Calibration plots.

Figure 3. Receiver operating characteristic curves for different modalities of CRC screening. CRC = colorectal cancer.

3.5. Validation of prediction model

The optimism is the difference between model performance in the bootstrap sample and in the original sample. In model 3 (Questionnaire + FIT), with 500 bootstrap replications, the estimated optimism was 0.00411 for EPV = 5, and the apparent performance for c-index of 0.8429 (= 0.8388 − 0.00411) shows good discrimination. With large sample size (EPV = 40 or 80), a reduction in optimism was found. The mean apparent c-index were 0.8438, 0.8407, 0.8372, and 0.8387 for EPV = 10, EPV = 20, EPV = 40, and EPV = 80, respectively. The 0.8387 of c-index for EPV = 80 was most close to the c-index of 0.8388 in original sample. Similar findings were observed in model 1 (Questionnaire only) and model 2 (FIT only). The mean apparent c-index were 0.7395 and 0.7315 for EPV = 5 and EPV = 80 in model 1. The mean apparent c-index were 0.7699 and 0.7639 for EPV = 5 and EPV = 80 in model 2.

4. Discussion

In contrast to previous risk prediction models that either incorporate dominant and high-penetrance genes into a genetic model or include personal characteristics and environmental risk factors in the non-genetic model, our risk prediction model for CRC is specific and considers the history of clinical symptoms and signs of bowel disease together with the FIT results. A risk score was developed after training regression coefficients as clinical weights to assess the influence that each correlate has on the risk for CRC by using empirical data from the QFIT program in Tianjin, China.

We found that the combination of information obtained from the questionnaire with FIT resulted in a good prediction for the risk of CRC, with an AUC up to 84% when using the non-parametric Mann–Whitney U method. The FIT information alone gives a lower prediction that is equivalent to that obtained

when using the questionnaire alone, with respect to AUC (76% vs 73%); however, this difference lacks statistical significance based on the Mann–Whitney U test.

The factors associated with CRC awareness could be considered predictors for predicting the risk of CRC. Such a prediction model might be useful for alerting someone with clinical symptoms or signs and thereby detecting CRC earlier.

The same logic is applied to administering a FIT to improve patients' awareness and detect early CRC cases. These clinical correlates may be a reflection of the proxy variables for residual familial risk, after making allowances for a family history of CRC.

Our prediction models were validated using bootstrapping method. The apparent performance for logistic regression model in data sets with 5 to 80 events per variable (EPV) with 500 bootstrap repetitions was estimated for internal validation. The low estimated optimism indicates a good discrimination for a given EPV value in our analysis. The apparent performance was more close to the performance in the total dataset while increasing the sample size (EPV = 80).

In light of these risk prediction models, providing an individual-guided risk approach is an alternative and efficient way to achieve the goal of mass screening for early detection of CRC in Asian countries that have a low or intermediate incidence rate. For example, the risk score can be built from data obtained in the first screening to develop a risk prediction model. This prediction model can then be applied in subsequent screenings by offering an individual-risk-guided invitation to a large population-based screening program.

The greatest strength of our risk prediction model is that it was developed by using large community- and population-based screening data. The threat to validity due to the selection bias that is inherent when using consecutive cases series from hospitals, as is usually found in case-control studies or clinical studies,[12] can

be ameliorated. Large community- and population-based studies have also gained sufficient statistical power for building a risk prediction model for CRC, as can be seen by the narrow confidence interval for the AUC of the ROC curves.

The main limitation is that we have not considered incorporating information into the genetic model about carrying a genetic mutation related to several major genes (such as the MMR genes, APC. MUTYH, STK11 [LB1], BMPR1A, and PTEN).[3] The binary property of family history is not adequate for capturing familial risk without considering the age of onset of CRC among relatives. Our risk prediction model is therefore not adequate for predicting CRC in association with familial risk. This should be considered when using family pedigree data obtained from, for example, a Keelung community-based integrated screening study.[27]

In conclusion, we have developed a CRC risk prediction model based on a series of symptoms and signs related to enteric diseases in combination with FIT. Such a risk prediction model is useful for improving the awareness of high-risk subjects and for individual-risk-guided invitation or strategies to achieve mass CRC screening.

Author contributions

Conceptualization: Li-Zhong Zhao, Dong-Wang Ma, Sam LI-SHENG Chen, Xi-Mo Wang.
Data curation: Dong-Wang Ma, De-Zheng Wang, Lei Shi, Hong-Lei Wang, Mo Dong, Shu-Yi Zhang, Lei Cao, Wei-Hua Zhang, Xi-Peng Zhang, Qing-Huai Zhang, Lin Yu, Hai Qin, Xi-Mo Wang.
Formal analysis: Li-Zhong Zhao, Sam LI-SHENG Chen.
Investigation: Li-Zhong Zhao, Dong-Wang Ma, De-Zheng Wang, Lei Shi, Hong-Lei Wang, Mo Dong, Shu-Yi Zhang, Lei Cao, Wei-Hua Zhang, Xi-Peng Zhang, Qing-Huai Zhang, Lin Yu, Hai Qin, Xi-Mo Wang.
Methodology: Li-Zhong Zhao, Sam LI-SHENG Chen.
Project administration: Dong-Wang Ma, Xi-Mo Wang.
Supervision: Xi-Mo Wang.
Validation: Sam LI-SHENG Chen.
Writing – original draft: Wen Li.
Writing – review & editing: Li-Zhong Zhao, Xi-Mo Wang.

References

[1] Win AK, Macinnis RJ, Hopper JL, et al. Risk prediction models for colorectal cancer: a review. Cancer Epidemiol Biomarkers Prev 2012;21:398–410.
[2] 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.
[3] Chen S, Wang W, Broman KW, et al. BayesMendel: an R environment for Mendelian risk prediction. Stat Appl Genet Mol Biol 2004;3:Article21.
[4] Chen S, Wang W, Lee S, et al. Prediction of germline mutations and cancer risk in the Lynch syndrome. JAMA 2006;296:1479–87.
[5] Palomaki GE, McClain MR, Melillo S, et al. EGAPP supplementary evidence review: DNA testing strategies aimed at reducing morbidity and mortality from Lynch syndrome. Genet Med 2009;11:42–65.
[6] von Holst S, Picelli S, Edler D, et al. Association studies on 11 published colorectal cancer risk loci. Br J Cancer 2010;103:575–80.
[7] Tenesa A, Dunlop MG. New insights into the aetiology of colorectal cancer from genome-wide association studies. Nat Rev Genet 2009;10:353–8.
[8] Antoniou AC, Pharoah PP, Smith P, et al. The BOADICEA model of genetic susceptibility to breast and ovarian cancer. Br J Cancer 2004;91:1580–90.
[9] Antoniou AC, Cunningham AP, Peto J, et al. The BOADICEA model of genetic susceptibility to breast and ovarian cancers: updates and extensions. Br J Cancer 2008;98:1457–66.
[10] Macinnis RJ, Antoniou AC, Eeles RA, et al. A risk prediction algorithm based on family history and common genetic variants: application to prostate cancer with potential clinical impact. Genet Epidemiol 2011;35:549–56.
[11] Colditz GA, Atwood KA, Emmons K, et al. Harvard report on cancer prevention volume 4: Harvard Cancer Risk Index. Cancer Causes Control 2000;11:477–88.
[12] Freedman AN, Slattery ML, Ballard-Barbash R, et al. Colorectal cancer risk prediction tool for white men and women without known susceptibility. J Clin Oncol 2009;27:686–93.
[13] Wei EK, Colditz GA, Giovannucci EL, et al. Cumulative risk of colon cancer up to age 70 years by risk factor status using data from the Nurses' Health Study. Am J Epidemiol 2009;170:863–72.
[14] Brenner H, Chang-Claude J, Seiler CM, et al. Long-term risk of colorectal cancer after negative colonoscopy. J Clin Oncol 2011;29:3761–7.
[15] Schoen RE. Surveillance after positive and negative colonoscopy examinations: issues, yields, and use. Am J Gastroenterol 2003;98:1237–46.
[16] Brenner H, Chang-Claude J, Seiler CM, et al. Protection from colorectal cancer after colonoscopy: a population based, case-control study. Ann Intern Med 2011;154:22–30.
[17] Imperiale TF, Wagner DR, Lin CY, et al. Using risk for advanced proximal colonic neoplasia to tailor endoscopic screening for colorectal cancer. Ann Intern Med 2003;139:959–65.
[18] Lin OS, Kozarek RA, Schembre DB, et al. Risk stratification for colon neoplasia: screening strategies using colonoscopy and computerized tomographic colonography. Gastroenterology 2006;131:1011–9.
[19] Driver JA, Gaziano JM, Gelber RP, et al. Development of a risk score for colorectal cancer in men. Am J Med 2007;120:257–63.
[20] Ma E, Sasazuki S, Iwasaki M, et al. 10-Year risk of colorectal cancer: development and validation of a prediction model in middle-aged Japanesemen. Cancer Epidemiol 2010;34:534–41.
[21] 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;60:1236–41.
[22] 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;175:584–93.
[23] Aaltonen L, Johns L, Järvinen H, et al. Explaining the familial colorectal cancer risk associated with mismatch repair (MMR)-deficient and MMR-stable tumors. Clin Cancer Res 2007;13:356–61.
[24] Chen LS, Yen AM, Chiu SY, et al. Baseline faecal occult blood concentration as a predictor of incident colorectal neoplasia: longitudinal follow-up of a Taiwanese population-based colorectal cancer screening cohort. Lancet Oncol 2011;6:551–68.
[25] Yen AM, Chen SL, Chiu SY, et al. A new insight into fecal hemoglobin concentration-dependent predictor for colorectal neoplasia. Int J Cancer 2014;135:1203–12.
[26] Peduzzi P, Concato J, Kemper E, et al. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996;49:1373–9.
[27] Chiu SY-H, Chen LS, Yen AM-F, et al. Population-based proband-oriented pedigree information system: application to hypertension with population-based screening data. J Am Med Inform Assoc 2012;19:102–10.