

A New Comprehensive Colorectal Cancer Risk Prediction Model Incorporating Family History, Personal Characteristics, and Environmental Factors



Yingye Zheng^{1,2}, Xinwei Hua^{1,3}, Aung K. Win^{4,5}, Robert J. MacInnis^{4,6}, Steven Gallinger⁷, Loic Le Marchand⁸, Noralane M. Lindor⁹, John A. Baron¹⁰, John L. Hopper⁴, James G. Dowty⁴, Antonis C. Antoniou¹¹, Jiayin Zheng¹, Mark A. Jenkins⁴, and Polly A. Newcomb^{1,3}

ABSTRACT

Purpose: Reducing colorectal cancer incidence and mortality through early detection would improve efficacy if targeted. We developed a colorectal cancer risk prediction model incorporating personal, family, genetic, and environmental risk factors to enhance prevention.

Methods: A familial risk profile (FRP) was calculated to summarize individuals' risk based on detailed cancer family history (FH), family structure, probabilities of mutation in major colorectal cancer susceptibility genes, and a polygenic component. We developed risk models, including individuals' FRP or binary colorectal cancer FH, and colorectal cancer risk factors collected at enrollment using population-based colorectal cancer cases ($N = 4,445$) and controls ($N = 3,967$) recruited by the Colon Cancer Family Registry Cohort (CCFRC). Model validation used CCFRC follow-up data for population-based ($N = 12,052$) and clinic-based ($N = 5,584$) relatives with no cancer history at recruitment to assess model calibration [expected/observed rate ratio (E/O)] and discrimination [area under the receiver-operating-characteristic curve (AUC)].

Results: The E/O [95% confidence interval (CI)] for FRP models for population-based relatives were 1.04 (0.74–1.45) for men and 0.86 (0.64–1.20) for women, and for clinic-based relatives were 1.15 (0.87–1.58) for men and 1.04 (0.76–1.45) for women. The age-adjusted AUCs (95% CI) for FRP models for population-based relatives were 0.69 (0.60–0.78) for men and 0.70 (0.62–0.77) for women, and for clinic-based relatives were 0.77 (0.69–0.84) for men and 0.68 (0.60–0.76) for women. The incremental values of AUC for FRP over FH models for population-based relatives were 0.08 (0.01–0.15) for men and 0.10 (0.04–0.16) for women, and for clinic-based relatives were 0.11 (0.05–0.17) for men and 0.11 (0.06–0.17) for women.

Conclusions: Both models calibrated well. The FRP-based model provided better risk stratification and risk discrimination than the FH-based model.

Impact: Our findings suggest detailed FH may be useful for targeted risk-based screening and clinical management.

Introduction

Screening for colorectal cancer is efficient and cost effective (1). The evidence is compelling, even when applied irrespective of personal characteristics except age (2). However, beyond age, individuals' colorectal cancer risk factors could inform the use and frequency of specific screening regimens (3, 4). A detailed risk prediction model would permit targeted screening at an appropriate level based on individuals' risk of colorectal cancer (5–8).

Family history (FH) of colorectal cancer is an important risk factor for this disease, as it is a consequence of genetic and environmental factors shared by relatives (9, 10). A comprehensive risk prediction model should incorporate detailed FH of cancer and available information on known genetic, and epidemiologic characteristics. To date, existing colorectal neoplasia prediction models included only limited information on FH, including simple measures of colorectal cancer FH or considering a group

¹Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington. ²Department of Biostatistics, University of Washington School of Public Health, Seattle, Washington. ³Department of Epidemiology, University of Washington School of Public Health, Seattle, Washington. ⁴Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Parkville, Victoria, Australia. ⁵Genetic Medicine, Royal Melbourne Hospital, Parkville, Victoria, Australia. ⁶Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Victoria, Australia. ⁷Lunenfeld Tanenbaum Research Institute, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada. ⁸Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii. ⁹Department of Health Science Research, Mayo Clinic Arizona, Scottsdale, Arizona. ¹⁰Department of Medicine, University of North Carolina School of Medicine, and Department of Epidemiology, Gillings School of Global Public Health, Chapel Hill, North Carolina. ¹¹Cancer Research UK, Genetic Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom.

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (<http://cebp.aacrjournals.org/>).

M.A. Jenkins and P.A. Newcomb are joint last authors of this article.

Prior presentation: Presented at the AACR Special Conference Improving Cancer Risk Prediction for Prevention and Early Detection; November 16–19, 2016; Orlando, FL. Abstract PR05: "Does a comprehensive family history of colorectal cancer improve risk prediction?"

Corresponding Author: Polly A. Newcomb, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N., M4-B402, Seattle, WA 98109. Phone: 206-667-3476; Fax: 206-667-7850; E-mail: pnewcomb@fredhutch.org

Cancer Epidemiol Biomarkers Prev 2020;29:549–57

doi: 10.1158/1055-9965.EPI-19-0929

©2020 American Association for Cancer Research.

of known low-penetrance genetic variants (11–15). Further, some published models included only limited risk factors; endoscopy is particularly important as it is one of the strongest risk factors for colorectal neoplasia (15–18). The performance of existing risk prediction models were inconsistently evaluated and described (18, 19). Models with external validation only gave reasonable discrimination, suggesting limited usefulness for risk-based screening (18–24).

The majority of these risk models defined FH as a binary variable, typically as “at least one first-degree relative with colorectal cancer”; a few models considered FH as the number of first- or second-degree relatives with colorectal cancer. A few models (e.g., MMRpro; ref. 25) used a more complex definition based on the number of affected relatives, their age at colorectal cancer diagnosis, and the degree of relatedness. In theory, the more detailed and accurate individual’s FH, the better the risk prediction. However, in a typical primary care setting with limited time and incomplete patient reports, only the presence or absence of a colorectal cancer FH is usually recorded. It is unclear whether such information is sufficient to predict colorectal cancer risk accurately. Using the large well-characterized data from the Colon Cancer Family Registry Cohort (CCFRC; refs. 26, 27), we describe the development and validation of a new risk prediction model that incorporates a novel measure of FH in addition to personal and environmental risk factors.

Materials and Methods

Study sample

The CCFRC is an NCI-funded international consortium of six colorectal cancer registries from the USA, Canada, and Australia/New Zealand, using standard protocols for data collection, molecular characterization, and follow-up at each site (<http://coloncfr.org/>; ref. 26). Recently diagnosed colorectal cancer cases from population-based cancer registries, controls from population-based sources (including drivers’ license, voting records, health beneficiary rosters, and electoral rolls), and cases from family cancer clinics with a strong FH of colorectal cancer or early-onset disease were recruited as “probands” between 1998 and 2012. Relatives of population- and clinic-based cases were also invited to participate. Written or verbal informed consent was obtained from all study participants. The institutional research ethics review boards at each study center approved the study protocols.

Data collection and genetic testing

At baseline, participants were asked to: (i) complete an epidemiologic risk factor questionnaire on medical history, demographic characteristics, reproductive history, physical activity, medication, postmenopausal hormone use, alcohol and tobacco use, and diet about 1 year before diagnosis or a comparable period in controls; (ii) describe detailed colorectal cancer FH information, at least for their first-degree relatives, including relationship to the participant, age, sex, and type and ages of cancer diagnosis; (iii) provide written consent for the research team to access tumor tissues and corresponding pathological reports; and (iv) collect a blood or buccal sample. Reported cancers and ages at diagnosis were confirmed, where possible, using pathology reports, medical records, cancer registry reports, and/or death certificates. Genetic mutation screening and testing for mismatch repair (MMR) genes (*MLH1*, *MSH2*, *MSH6*, *PMS2*) and *MUTYH* was completed for probands and relatives, as previously described in detail (28). During follow-up

approximately every 5 years, participants from all case families were contacted for updates on incident polyps, cancer diagnoses at any site, colorectal cancer screening and surgery, and their relatives’ cancer diagnoses and deaths. Population-based controls were not followed up.

Family history measures

FH of colorectal cancer was defined in two ways: (i) as a binary indicator (yes/no) of having at least one first-degree relative with colorectal cancer (hereafter called, FH); and (ii) as familial risk profile (FRP), a continuous measure indicating absolute risk of colorectal cancer from birth to age 80 years. The FRP was calculated based on detailed cancer FH, considering age at diagnosis as well as the number of and relationship to each relative, their ages, their probabilities of carrying colorectal cancer predisposing mutations in the DNA MMR genes (*MLH1*, *MSH2*, *MSH6*, *PMS2*) and *MUTYH* (28, 29), and the residual familial aggregation of colorectal cancer risk not explained by the known major genes. We performed modified segregation analysis and used a mixed model to account for the hypothetical unidentified major genes by fitting an unmeasured polygene and a dominant major gene component. The unidentified major genes were autosomal with a normal and a mutant allele unlinked to mutations in the MMR genes or *MUTYH*. The polygenic component was approximated by the hypergeometric polygenic model (30, 31) and was assumed to be normally distributed and age dependent. Specifically, the calculation of FRP uses: (i) the age-, sex-, and country-specific incidence of colorectal cancer from national cancer statistics; (ii) the familial relative risk based on previous segregation analysis of colorectal cancer data from the CCFRC (28); and (iii) the age-specific incidence of colorectal cancer based on mutation status, for which we used the penetrance reported from analyses of the CCFRC (Supplementary Methods; refs. 28, 32–36).

Model development

We included 4,445 colorectal cancer cases and 3,967 controls recruited from three population-based sites of the CCFRC (Seattle, WA, USA; Ontario, Canada; and Victoria, Australia) for model development. Only cases who were diagnosed with colorectal cancer less than 2 years before completing the baseline questionnaire were used to ensure that all risk factors pertained to the prediagnostic period. Controls were frequency-matched to the cases on age. We restricted our analysis to non-Hispanic whites. Participants with missing values on any of the baseline variables were excluded from model development.

We included established prediagnosis risk factors in model development, including screening history. A list of candidate variables and their parameterization used for model selection is described in Supplementary Table S1. The colorectal cancer risk prediction model was developed using logistic regression with case–control status as the outcome. The distributions of FRP by sex and case–control status were examined using histograms and compared using Wilcoxon nonparametric tests. Models using either log-transformed FRP or FH were stratified by sex to permit sex-specific associations with risk factors.

Three forward-stepwise-selection procedures were implemented. Each used a different selection criterion: (i) P value < 0.15 ; (ii) incremental value in AUC ($\text{incAUC} \geq 0.01$); or (iii) a smaller Akaike information criterion (AIC; refs. 37, 38). Final models from these different selection criteria were compared to identify variables that were robust to selection procedures. For example, if smoking pack years was included in the variable set in a model based on incAUC , and as well as that from an AIC-based selection procedure, then we

Table 1. Associations between risk factor variables and colorectal cancer from the final risk prediction model with familial risk profile (FRP model) and the final risk prediction model with a binary family history (FH model), for men only.

Variables	Cases (N = 2,312) ^a N (%) or mean (SD)	Controls (N = 1,916) ^a N (%) or mean (SD)	FRP model ^b OR (95% CI)	FH model ^b OR (95% CI)
Family history				
FRP, mean (SD)	0.09 (0.097)	0.07 (0.027)	1.16 (1.11–1.20) ^c	—
Binary FH, N (%)				
No	1,859 (80.4)	1,731 (90.3)	—	1.00 (Ref)
Yes	453 (19.6)	185 (9.7)		2.34 (1.90–2.88)
Recent BMI ^d , kg/m ²				
<25	571 (24.6)	600 (31.3)	1.00 (Ref)	1.00 (Ref)
25–30	1,113 (48.0)	940 (49.1)	1.38 (1.18–1.62)	1.35 (1.15–1.58)
>30	611 (26.3)	372 (19.4)	1.59 (1.31–1.93)	1.61 (1.32–1.95)
Red meat consumption, servings/day				
<1	1,681 (72.4)	1,484 (77.5)	1.00 (Ref)	1.00 (Ref)
1+	564 (24.3)	364 (19.0)	1.27 (1.08–1.51)	1.25 (1.06–1.47)
Regular NSAID use duration ^e , years				
Nonuser	1,364 (58.8)	939 (49.0)	1.00 (Ref)	1.00 (Ref)
≤2	475 (20.5)	388 (20.3)	0.91 (0.76–1.09)	0.93 (0.78–1.10)
>2	406 (17.5)	509 (26.6)	0.73 (0.61–0.87)	0.72 (0.61–0.86)
Calcium supplement use duration, years				
Nonuser	2,060 (88.8)	1,610 (84.0)	1.00 (Ref)	1.00 (Ref)
≤2.5	115 (5.0)	97 (5.1)	1.29 (0.95–1.75)	1.32 (0.97–1.79)
>2.5	84 (3.6)	137 (7.2)	0.62 (0.46–0.83)	0.62 (0.46–0.83)
Cigarette smoking, pack-years				
Never	810 (34.9)	689 (36.0)	1.00 (Ref)	1.00 (Ref)
<10	332 (14.3)	281 (14.7)	1.00 (0.81–1.23)	1.03 (0.84–1.27)
10–19	289 (12.5)	229 (12.0)	1.18 (0.95–1.47)	1.16 (0.93–1.45)
20+	807 (34.8)	632 (33.0)	1.31 (1.11–1.55)	1.34 (1.14–1.58)
History of polyp ^f				
No	2,079 (89.6)	1,691 (88.3)	1.00 (Ref)	1.00 (Ref)
Yes	159 (6.9)	197 (10.3)	1.46 (1.10–1.95)	1.46 (1.10–1.94)
History of FOBT ^f				
No	1,569 (67.6)	1,175 (61.3)	1.00 (Ref)	1.00 (Ref)
Yes	646 (27.8)	667 (34.8)	0.79 (0.67–0.93)	0.80 (0.68–0.95)
History of sigmoidoscopy ^f				
No	1,893 (81.6)	1,397 (72.9)	1.00 (Ref)	1.00 (Ref)
Yes	330 (14.2)	434 (22.7)	0.86 (0.71–1.05)	0.85 (0.70–1.03)
History of colonoscopy ^f				
No	2,035 (87.7)	1,536 (80.2)	1.00 (Ref)	1.00 (Ref)
Yes	217 (9.3)	332 (17.3)	0.47 (0.37–0.60)	0.48 (0.38–0.61)

Abbreviations: BMI, body mass index; CI, confidence interval; FH, family history; FOBT, fecal occult blood test; FRP, familial risk profile; OR, odds ratio.

^aDo not sum to total due to missing values; N for each category of variables unless otherwise specified.

^bBoth models adjusted for study site and age at reference (age at diagnosis for cases and age at baseline interview for controls) in addition to the variables presented in this table.

^cOR for FRP: per 10% increase.

^dAs of 2 years before enrollment.

^eRegular NSAID use was defined as use of aspirin and/or ibuprofen at least twice a week for more than a month.

^fHistory of polyp, FOBT, sigmoidoscopy, and colonoscopy was defined as history of having any of these conditions/tests 2 years before enrollment.

would include it as a predictor in the final model even if it did not pass the *P* value threshold. In addition to the final list of personal characteristics and environmental factors, we included two definitions of FH (FH and log-transformed FRP) in our final models and further adjusted for study site and reference age (age at diagnosis of colorectal cancer for cases and age at interview for controls) in all models. To account for site-specific sampling of CCFRC, we applied weights based on the (inverse) sampling probability of each individual.

Relative and absolute risk calculation

ORs and corresponding 95% confidence intervals (CI) were generated from the final models. To calculate absolute risk, we

obtained sex- and age-specific colorectal cancer incidences for the U.S. (SEER-9 Registries, whites), Australia (Victoria), and Canada (Ontario) populations from the Cancer Incidence in Five Continents (CI5), for 1998 to 2002 (36), corresponding approximately to the time period during which cases were diagnosed. Deaths from non-colorectal cancer causes were considered as competing risks. Age- and sex-specific mortality from causes other than colorectal cancer was obtained from all-cause and colorectal cancer-specific mortality for the United States, Australia, and Canada, respectively during the same time period (39–42). Five-year absolute risks were calculated as described in Freedman and colleagues (ref. 43; Supplementary Methods).

Model validation

For model validation, we studied 17,636 unaffected relatives of case probands who: (i) were recruited from population registries ($N = 12,052$) and genetic clinics ($N = 5,584$); (ii) were non-Hispanic whites; (iii) had no personal history of any cancer at the time of recruitment; (iv) completed a baseline questionnaire; (v) were recruited from five study sites of the CCFRC (Australia/New Zealand, Mayo Clinic, Ontario, Cedars-Sinai, and Seattle); and (vi) were prospectively followed up after the baseline recruit-

ment (27). The flowchart of the study design and model steps is included as Supplementary Fig. S1.

Colorectal cancer incidence rates are higher in the clinical compared with the general population. Therefore, we assessed the model calibration separately using different baseline risks. For population-based set, the baseline risks were derived from the age-, sex-, and country-specific incidence rates from the general population (36). For clinical set with relatively higher colorectal cancer incidence, we calculated the baseline risks using the clinic-based validation set. We then calculated

Table 2. Associations between risk factor variables and colorectal cancer from the final risk prediction model with familial risk profile (FRP model) and from the final risk prediction model with a binary family history (FH model), for women only.

Variables	Cases ($N = 2,133$) ^a N (%) or mean (SD)	Controls ($N = 2,051$) ^a N (%) or mean (SD)	FRP model ^b OR (95% CI)	FH model ^b OR (95% CI)
Family history				
FRP, mean (SD)	0.07 (0.090)	0.05 (0.024)	1.09 (1.06–1.12) ^c	—
Binary FH, N (%)				
No	1,708 (80.1)	1,799 (87.7)	—	1.00 (Ref)
Yes	425 (19.9)	252 (12.3)		1.72 (1.39–2.12)
Recent BMI ^d , kg/m ²				
<25	971 (45.5)	1,055 (51.4)	1.00 (Ref)	1.00 (Ref)
25–30	631 (29.6)	591 (28.8)	1.19 (1.00–1.41)	1.20 (1.02–1.43)
>30	501 (23.5)	380 (18.5)	1.40 (1.15–1.70)	1.42 (1.17–1.72)
Red meat consumption, servings/day				
<1	1,721 (80.7)	1,694 (82.6)	1.00 (Ref)	1.00 (Ref)
1+	288 (13.5)	228 (11.1)	1.48 (1.18–1.85)	1.47 (1.17–1.85)
Fruit consumption, servings/day				
<1	525 (24.6)	427 (20.8)	1.00 (Ref)	1.00 (Ref)
1+	1,536 (72.0)	1,588 (77.4)	0.83 (0.70–0.99)	0.83 (0.69–0.99)
Smoking, pack-years				
Never	1,003 (47.0)	1,047 (51.0)	1.00 (Ref)	1.00 (Ref)
<10	392 (18.4)	363 (17.7)	1.13 (0.93–1.38)	1.16 (0.95–1.41)
10 to 19	236 (11.1)	174 (8.5)	1.13 (0.88–1.45)	1.13 (0.88–1.45)
20 +	411 (19.3)	388 (18.9)	1.02 (0.84–1.25)	1.04 (0.85–1.26)
Calcium supplement use duration, years				
Nonuser	1,268 (59.4)	985 (48.0)	1.00 (Ref)	1.00 (Ref)
≤2.5	294 (13.8)	279 (13.6)	0.98 (0.80–1.21)	0.96 (0.78–1.19)
>2.5	378 (17.7)	564 (27.5)	0.81 (0.67–0.97)	0.80 (0.67–0.96)
History of polyp ^e				
No	1,943 (91.1)	1,835 (89.5)	1.00 (Ref)	1.00 (Ref)
Yes	129 (6.0)	177 (8.6)	1.48 (1.06–2.07)	1.50 (1.07–2.09)
History of FOBT ^e				
No	1,502 (70.4)	1,201 (58.6)	1.00 (Ref)	1.00 (Ref)
Yes	541 (25.4)	793 (38.7)	0.78 (0.65–0.94)	0.77 (0.64–0.93)
History of sigmoidoscopy ^e				
No	1,747 (81.9)	1,472 (71.8)	1.00 (Ref)	1.00 (Ref)
Yes	281 (13.2)	502 (24.5)	0.68 (0.56–0.84)	0.68 (0.55–0.84)
History of colonoscopy ^e				
No	1,890 (88.6)	1,635 (79.7)	1.00 (Ref)	1.00 (Ref)
Yes	197 (9.2)	373 (18.2)	0.39 (0.30–0.51)	0.39 (0.30–0.52)
Postmenopausal hormone use				
Nonuser	1,591 (74.6)	1,347 (65.7)	1.00 (Ref)	1.00 (Ref)
Estrogen only	207 (9.7)	277 (13.5)	1.03 (0.82–1.31)	1.05 (0.83–1.33)
Estrogen + progesterone only	126 (5.9)	183 (8.9)	0.81 (0.61–1.07)	0.82 (0.62–1.08)
Mixed	96 (4.5)	145 (7.1)	0.92 (0.67–1.27)	0.90 (0.66–1.24)

Abbreviations: BMI, body mass index; CI, confidence interval; FH, family history; FOBT, fecal occult blood test; FRP, familial risk profile; OR, odds ratio.

^aDo not sum to total due to missing values; N for each categories of variables unless otherwise specified.

^bBoth models adjusted for study site and age at reference (age at diagnosis for cases and age at baseline interview for controls) in addition to the variables presented in this table.

^cOR for FRP: per 10% increase.

^dAs of 2 years before enrollment.

^eHistory of polyp, FOBT, sigmoidoscopy, and colonoscopy was defined as history of having any of these conditions/tests 2 years before enrollment.

absolute risk for each individual in the validation set based on the relative risks from our final model, the derived baseline risks, and risk factor data from the baseline questionnaire. In total, 12.6% of men and 17.0% of women had at least one covariate with missing data (see Supplementary Table S2). We conducted imputation for each risk factor variable in the final model with the most frequent (modal) category for men and women separately. Model performance was compared before and after exclusion of relatives with missing covariates (Supplementary Table S3). Using the follow-up data of the CCFRC, we identified incident colorectal cancer diagnoses. Time to event was defined as years from the date of baseline interview completion to the date of diagnosis of incident colorectal cancer. Individuals with a diagnosis of other types of cancer were censored at the date of diagnosis. Deaths due to causes other than colorectal cancer were considered as competing risk events. Participants who were alive without any cancer diagnosis were censored at the date of last contact.

We then categorized participants into quintiles of predicted absolute risk based on the developed model and plotted the average observed absolute risk within each quintile against the predicted risks in that quintile, adding 95% CIs of the observed risks. The observed marginal risks were calculated as the cumulative incidences of colorectal cancer, accounting for censoring and competing risks. In addition, we calculated a summary measure of calibration for the FRP and FH models as the ratio of the averaged predicted 5-year absolute risk to the observed cumulative incidence rate (E/O), separately for men and women; 95% CIs were calculated using a bootstrap approach (44). To assess the performance of the model for risk stratification, we defined four risk groups based on the 3rd, 6th, and 9th deciles of the predicted 5-year absolute risks of colorectal cancer. For each model, we plotted cumulative incidence functions of colorectal cancer diagnosis by risk groups, as above, and tested differences in risk functions across groups using K-sample test (45).

For discrimination performance, we used ROC curve analyses and calculated AUC (Supplementary Methods) to assess the model's ability to separate individuals with and without a colorectal cancer diagnosis within 5 years after baseline. Since the outcome of individuals who were censored within 5 years after baseline was uncertain, we excluded these individuals in this calculation and used inverse censoring probability weights to account for the missing information. Two age groups were defined as ≤ 50 and >50 years old at baseline separately for both men and women. Age-adjusted ROC curves and AUCs were calculated as the weighted average of age-specific AUC, with weights as the proportion of colorectal cancer diagnosis in each age group. We used a bootstrap approach to calculate the empirical 95% CIs for age-adjusted AUC based on 2.5th and 97.5th percentiles. Analyses were conducted using R version 3.1.1 (<http://www.r-project.org/>). All statistical tests were two-sided, and *P* values of less than 0.05 were considered statistically significant.

Results

Model development

Using population-based cases and controls, the variables entered into the final models of FRP and FH for men and women are shown in **Tables 1** and **2**, respectively. The distribution of FRP (range, 0.037–0.993) by sex and case–control status is summarized in Supplementary Table S4 and Supplementary Fig. S2. For both men and women, cases had higher FRP than controls (all *P* < 0.001).

For men, every 10% relative increase in FRP (e.g., 0.33 vs. 0.30) was associated with 16% higher risk of developing colorectal cancer (95% CI, 11%–20%). From the FH model, the OR for FH was 2.34 (95% CI,

Table 3. Observed 5-year cumulative incidence rates (O) versus averaged 5-year absolute risk (E) based on risk models with FRP or with a binary family history, and separately for men and women.

Sex/Model	Expected averaged 5-year absolute risk (per 1,000 person-years; 95% CI)	Observed 5-year cumulative incidence rates (per 1,000 person-years; 95% CI)	E/O (95% CI)
<i>Population-based relatives</i>			
Men			
FRP model	7.1 (6.7–7.5)	7.1 (4.9–9.6)	1.04 (0.74–1.45)
FH model	6.1 (5.9–6.3)		
Women			
FRP model	5.2 (5.0–5.5)	6.3 (4.4–8.2)	0.86 (0.64–1.20)
FH model	5.2 (5.0–5.4)		
<i>Clinic-based relatives</i>			
Men			
FRP model	21.2 (19.9–22.5)	18.9 (13.3–24.4)	1.15 (0.87–1.58)
FH model	19.8 (19.2–20.5)		
Women			
FRP model	13.6 (13.0–14.2)	13.4 (9.3–17.8)	1.04 (0.76–1.45)
FH model	13.2 (12.8–13.7)		

1.90–2.88). The strengths of association with the other variables were similar for FRP and FH models (**Table 1**). For women, a 10% relative increase in FRP was associated with 9% higher risk of colorectal cancer (95% CI, 6%–12%). From the FH model, the OR for FH was 1.72 (95% CI, 1.39–2.12). The strengths of associations with other variables were essentially no different than those from the FRP model (**Table 2**).

Model validation

The median follow-up time was 8.6 years; 317 relatives were diagnosed with incident colorectal cancer during this period. Calibration for population- and clinic-based relatives across different risk groups is presented in Supplementary Fig. S3. The overall E/O estimates (95% CI) for different models are summarized in **Table 3**. For population-based relatives, FRP and FH models calibrated well, with E/O estimates (95% CI) of 1.0 (0.7–1.4) and 0.9 (0.6–1.2) for men and women from FRP models, and 0.9 (0.6–1.2) and 0.8 (0.6–1.2) from FH models. For clinic-based relatives, FRP and FH models calibrated well with E/O (95% CI) ranging from 1.0 (0.8–1.4) to 1.2 (0.9–1.6).

In addition, we defined four groups at different levels of predicted risks (using 3rd, 6th, and 9th deciles as cutoffs). The cumulative incidence curves are presented for population- and clinic-based relatives separately (**Fig. 1A** and **B**). The wider separation of the FRP models suggests they performed better than the FH models in stratifying individuals into distinctive risk groups.

The FRP model also provided improved discriminatory capacity over the FH model (**Fig. 2**). For population-based relatives, the age-adjusted AUCs (95% CI) for the FRP model were 0.69 (0.60–0.78) for men and 0.70 (0.62–0.77) for women (**Table 4**). The increments in age-adjusted AUCs [incAUC (95% CI)] for FRP over FH models were 0.08 (0.01–0.15) for men and 0.10 (0.04–0.16) for women (**Table 4**). For clinic-based relatives, the age-adjusted AUCs (95% CI) for FRP models were 0.77 (0.69–0.84) and 0.68 (0.60–0.76) for men and women, respectively. The incAUCs (95% CI) for FRP over FH models was 0.11 (0.05–0.17) for men and 0.11 (0.06–0.17) for women.

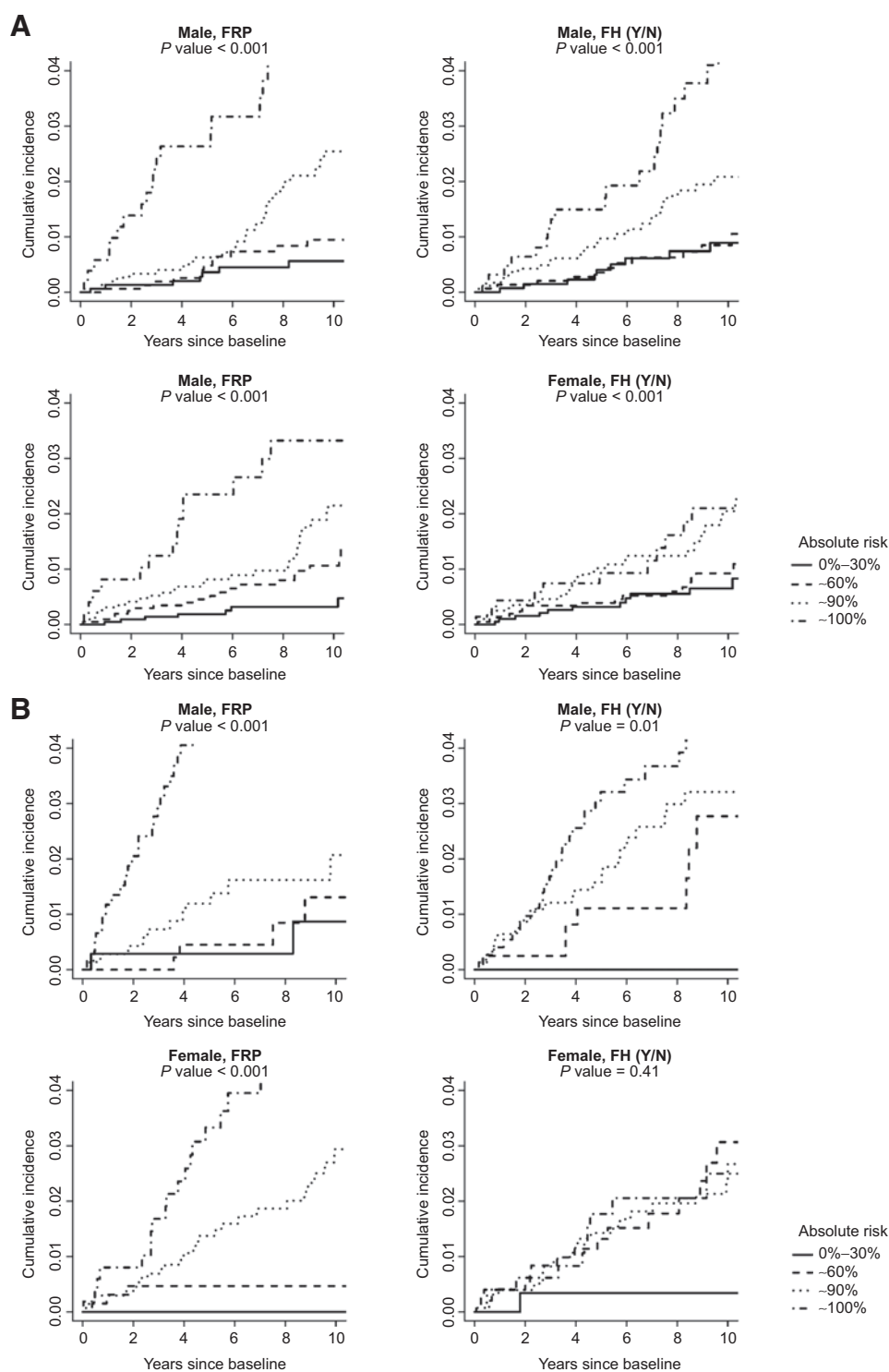


Figure 1. Cumulative incidence of colorectal cancer according to estimated 5-year absolute risk for population-based relatives (A) and clinic-based relatives (B). Four groups were defined based on cutpoints of 3rd, 6th, and 9th deciles of estimated 5-year absolute risk. The K-sample test was used to compare the cumulative incidence across groups and to calculate two-sided P values (45).

Downloaded from <http://aacrjournals.org/cebp/article-pdf/29/3/549/2340006/549.pdf> by guest on 10 August 2024

Discussion

We developed and validated a new risk prediction model which incorporated detailed FH information captured by the FRP as well as personal and environmental risk factors. Generally, both FRP and FH models provided good calibration; however, our results found that the FRP-based model gave better discrimination than a model using a

simple binary summary of FH (this model is available online at <http://crisptool.org/crisp-int>). Furthermore, our results suggest that the model that incorporated FRP is more valuable for risk stratification, giving distinct risk categories for men and women.

One clinical utility of colorectal cancer risk models is to provide information for screening regimens tailored to an individual's risk, and to inform intensity of screening, decisions regarding

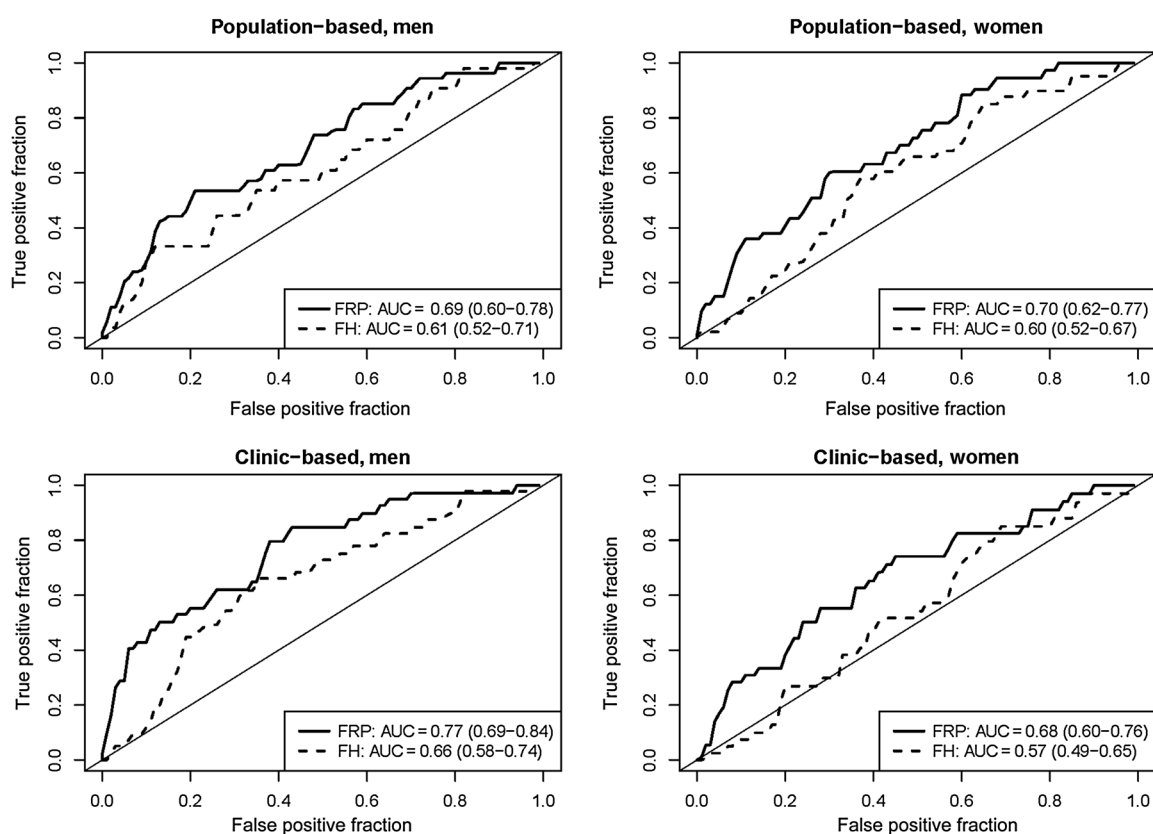


Figure 2.

Age-adjusted ROC curves for men and women. ROC curves and age-adjusted AUC were calculated as the weighted average of age-specific estimates, with weights as the proportion of colorectal cancer diagnosis in each age group (<50 and \geq 50 at baseline). We calculated 95% CIs (in parentheses) using bootstrap approach.

chemoprevention, and utilization of gene panel testing. Current colorectal cancer screening recommendations are based solely on age and simple measures of FH. Our study suggests that consideration of multiple risk factors, including a detailed FH of colorectal cancer, can lead to the identification of individuals across the spectrum of colorectal cancer risk, from those at very low risk for whom delayed and/or noninvasive screening strategies are appropriate to those at high risk for whom earlier screening and more frequent/invasive monitoring is recommended. We have shown that FH of colorectal cancer is an important factor for colorectal cancer risk prediction, either defined as a binary (yes/no) measure or based on FRP calculated from the family structure, cancer histories, and MMR/*MUTYH* mutation status. Our research supports two approaches to risk prediction for colorectal cancer. In settings where FH information is limited, the risk model could include only the simple present/absent question. In situations such as genetic clinics where FH information is likely to be more complete, the risk model could make use of the FRP to derive more precise risk discrimination.

Numerous risk models have been developed to predict colorectal cancer and colorectal adenomas based on colorectal cancer FH, genetic mutation screening, SNP testing, personal characteristics, and known risk factors—singly or in combination (7, 18, 19, 24, 46). However, our FRP-based risk model is unique in its incorporation of all these colorectal cancer risk factors and in its use of our novel familial risk measure based on detailed FH information. Risk models that use FH as

a binary indicator do not account for variability in family size, age, or structure, age of colorectal cancer diagnosis, or the relationship of affected relatives to the proband, which are integral to characterizing familial risk (9, 29, 47). Both models evaluated in our study included environmental factors, as have most prediction models, to take advantage of the substantial contribution of these exposures on colorectal cancer risk (48).

Our study has many strengths, including its large study population where the colorectal cancer cases had a broad spectrum of familial risk. All risk factors were collected using a standardized instrument by the CCFRC sites. In particular, the assessment of family structure and cancer history was extensive (29, 47). The effect sizes of the general epidemiologic factors were largely consistent with previous

Table 4. Comparison of model performance using age-adjusted AUC (95% CI) by sex, by population- versus clinic-based families.

Models	FRP model	FH model	incAUC
Population-based			
Male	0.69 (0.60-0.78)	0.61 (0.52-0.71)	0.08 (0.01-0.15)
Female	0.70 (0.62-0.77)	0.60 (0.52-0.67)	0.10 (0.04-0.16)
Clinic-based			
Male	0.77 (0.69-0.84)	0.66 (0.58-0.74)	0.11 (0.05-0.17)
Female	0.68 (0.60-0.76)	0.57 (0.49-0.65)	0.11 (0.06-0.17)

literature (43, 49). The broad ascertainment of risk factors permitted the inclusion of several risk factors not in earlier models, such as endoscopy history. Finally, the CCFRC's use of a prospective follow-up design provided a validation data set with the same well-annotated information and from the same families upon which the model was developed. It would be interesting to evaluate the calibration of this model in an external validation set, although a challenge to identify a study with well-annotated FH.

Our study had some limitations. Although the validation of these models was prospective, with epidemiologic factors assessed at baseline, its development was based on retrospective reports of exposures prior to recruitment. Better assessment of relevant exposures could improve the predictive ability of these models. In addition, since our cohort started well over a decade ago, information on colorectal cancer screening might not reflect current screening practices. Our model could not be directly compared with other published prediction models due to differences in population structure, covariate ascertainment, and our statistical approach (50). However, our model development yielded similar covariates as the well-accepted Freedman prediction model (43). It is reasonable to consider our comparison (between the FRP vs. FH model) is an approximate comparison of our comprehensive FRP model with the Freedman model, and performed well. Finally, the FRP-based model could possibly be enhanced by including susceptibility genetic variants and other newly identified epigenetic changes (51).

In conclusion, we developed and validated a new colorectal cancer risk prediction model that incorporates a novel FH measure, the FRP, in addition to personal characteristics and other non-FH-based risk factors. The new FRP model provided better risk discrimination than the FH model, suggesting that more detailed FH has the potential to be informative for risk-based clinical decision-making.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Disclaimer

The funders had no role in the design of the study; the collection, analysis, and interpretation of the data; the writing of the article; and the decision to submit the article for publication.

References

1. Doubeni CA. Precision screening for colorectal cancer: promise and challenges. *Ann Intern Med* 2015;163:390–1.
2. Slattery ML, Edwards SL, Ma KN, Friedman GD. Colon cancer screening, lifestyle, and risk of colon cancer. *Cancer Causes Control* 2000;11:555–63.
3. Hagggar FA, Boushey RP. Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. *Clin Colon Rectal Surg* 2009;22:191–7.
4. Potter JD. Risk factors for colon neoplasia—epidemiology and biology. *Eur J Cancer* 1995;31A:1033–8.
5. Freedman AN, Seminara D, Gail MH, Hartge P, Colditz GA, Ballard-Barbash R, et al. Cancer risk prediction models: a workshop on development, evaluation, and application. *J Natl Cancer Inst* 2005;97:715–23.
6. Zauber AG, Lansdorp-Vogelaar I, Knudsen AB, Wilschut J, van Ballegoijen M, Kuntz KM. Evaluating test strategies for colorectal cancer screening: a decision analysis for the U.S. Preventive Services Task Force. *Ann Intern Med* 2008;149:659–69.
7. Win AK, MacInnis RJ, Hopper JL, Jenkins MA. Risk prediction models for colorectal cancer: a review. *Cancer Epidemiol Biomarkers Prev* 2012;21:398–410.
8. Taplin SH, Ichikawa L, Buist DS, Seger D, White E. Evaluating organized breast cancer screening implementation: the prevention of late-stage disease? *Cancer Epidemiol Biomarkers Prev* 2004;13:225–34.
9. Taylor DP, Burt RW, Williams MS, Haug PJ, Cannon-Albright LA. Population-based family history-specific risks for colorectal cancer: a constellation approach. *Gastroenterology* 2010;138:877–85.
10. Henrikson NB, Webber EM, Goddard KA, Scrol A, Piper M, Williams MS, et al. Family history and the natural history of colorectal cancer: systematic review. *Genet Med* 2015;17:702–12.
11. Jenkins MA, Makalic E, Dowty JG, Dite GS, MacInnis RJ, Ait Ouakrim D, et al. Quantifying the utility of single nucleotide polymorphisms to guide colorectal cancer screening. *Future Oncol* 2016;12:503–13.
12. Weigl K, Thomsen H, Balavarca Y, Hellwege JN, Shrubsole MJ, Brenner H. Genetic risk score is associated with prevalence of advanced neoplasms in a colorectal cancer screening population. *Gastroenterology* 2018;155:88–98.e10.
13. Weigl K, Chang-Claude J, Knebel P, Hsu L, Hoffmeister M, Brenner H. Strongly enhanced colorectal cancer risk stratification by combining family history and genetic risk score. *Clin Epidemiol* 2018;10:143–52.

Authors' Contributions

Conception and design: Y. Zheng, X. Hua, A.K. Win, R.J. MacInnis, J.L. Hopper, M.A. Jenkins, P.A. Newcomb

Development of methodology: Y. Zheng, A.K. Win, R.J. MacInnis, J.L. Hopper, J.G. Dowty, A.C. Antoniou, M.A. Jenkins, P.A. Newcomb

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): A.K. Win, S. Gallinger, L. Le Marchand, N.M. Lindor, J.A. Baron, J.L. Hopper, M.A. Jenkins, P.A. Newcomb

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): Y. Zheng, X. Hua, A.K. Win, R.J. MacInnis, J.L. Hopper, J.G. Dowty, J. Zheng, M.A. Jenkins, P.A. Newcomb

Writing, review, and/or revision of the manuscript: Y. Zheng, X. Hua, A.K. Win, R.J. MacInnis, L. Le Marchand, N.M. Lindor, J.A. Baron, J.L. Hopper, J.G. Dowty, A.C. Antoniou, M.A. Jenkins, P.A. Newcomb

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): S. Gallinger, J.L. Hopper, P.A. Newcomb

Study supervision: M.A. Jenkins, P.A. Newcomb

Acknowledgments

The authors thank all study participants of the Colon Cancer Family Registry and staff for their many contributions to this project.

This work was supported by R01 CA170122 from the NCI, NIH, and through cooperative agreements with the following Colon Cancer Family Registry (CCFR) centers: Australasian Colorectal Cancer Family Registry (U01/U24 CA097735), Mayo Clinic Cooperative Family Registry for Colon Cancer Studies (U01/U24 CA074800), Ontario Familial Colorectal Cancer Registry (U01/U24 CA074783), Seattle Colorectal Cancer Family Registry (U01/U24 CA074794), and USC Consortium Colorectal Cancer Family Registry (U01/U24 CA074799).

Seattle CCFR research was also supported by the Seattle-Puget Sound Surveillance Epidemiology and End Results (SEER) registry, which was funded by control nos. N01-CN-67009 and N01-PC-35142 and contract no. HHSN2612013000121 from the SEER Program of the NCI. Additional support included grants from the NIH UM1/U01 CA167551, K05 CA152715, and R01 CA236558, and through the Centre for Research Excellence grant APP1042021 and Program Grant APP1074383 from the National Health and Medical Research Council (NHMRC), Australia.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received August 23, 2019; revised October 29, 2019; accepted January 7, 2020; published first January 13, 2020.

14. McGeoch L, Saunders CL, Griffin SJ, Emery JD, Walter FM, Thompson DJ, et al. Risk prediction models for colorectal cancer incorporating common genetic variants: a systematic review. *Cancer Epidemiol Biomarkers Prev* 2019;28:1580–93.
15. Taylor DP, Stoddard GJ, Burt RW, Williams MS, Mitchell JA, Haug PJ, et al. How well does family history predict who will get colorectal cancer? Implications for cancer screening and counseling. *Genet Med* 2011;13:385–91.
16. Shin A, Joo J, Yang HR, Bak J, Park Y, Kim J, et al. Risk prediction model for colorectal cancer: National Health Insurance Corporation Study, Korea. *PLoS One* 2014;9:e88079.
17. Ma E, Sasazuki S, Iwasaki M, Sawada N, Inoue M, Tsugane S, et al. 10-Year risk of colorectal cancer: development and validation of a prediction model in middle-aged Japanese men. *Cancer Epidemiol* 2010;34:534–41.
18. Peng L, Weigl K, Boakye D, Brenner H. Risk scores for predicting advanced colorectal neoplasia in the average-risk population: a systematic review and meta-analysis. *Am J Gastroenterol* 2018;113:1788–800.
19. Smith T, Muller DC, Moons KGM, Cross AJ, Johansson M, Ferrari P, et al. Comparison of prognostic models to predict the occurrence of colorectal cancer in asymptomatic individuals: a systematic literature review and external validation in the EPIC and UK Biobank prospective cohort studies. *Gut* 2019;68:672–83.
20. Yeoh KG, Ho KY, Chiu HM, Zhu F, Ching JY, Wu DC, 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.
21. 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;12:478–85.
22. Kim DH, Cha JM, Shin HP, Joo KR, Lee JJ, Park DI. Development and validation of a risk stratification-based screening model for predicting colorectal advanced neoplasia in Korea. *J Clin Gastroenterol* 2015;49:41–9.
23. Jung YS, Park CH, Kim NH, Lee MY, Park DI. Impact of age on the risk of advanced colorectal neoplasia in a young population: an analysis using the predicted probability model. *Dig Dis Sci* 2017;62:2518–25.
24. Usher-Smith JA, Walter FM, Emery JD, Win AK, Griffin SJ. Risk prediction models for colorectal cancer: a systematic review. *Cancer Prev Res (Phila)* 2016;9:13–26.
25. Chen S, Wang W, Lee S, Nafa K, Lee J, Romans K, et al. Prediction of germline mutations and cancer risk in the Lynch syndrome. *JAMA* 2006;296:1479–87.
26. Newcomb PA, Baron J, Cotterchio M, Gallinger S, Grove J, Haile R, et al. Colon Cancer Family Registry: an international resource for studies of the genetic epidemiology of colon cancer. *Cancer Epidemiol Biomarkers Prev* 2007;16:2331–43.
27. Jenkins MA, Win AK, Templeton AS, Angelakos MS, Buchanan DD, Cotterchio M, et al. Cohort profile: The Colon Cancer Family Registry Cohort (CCFRC). *Int J Epidemiol* 2018;47:387–8i.
28. Win AK, Jenkins MA, Dowty JG, Antoniou AC, Lee A, Giles GG, et al. Prevalence and penetrance of major genes and polygenes for colorectal cancer. *Cancer Epidemiol Biomarkers Prev* 2017;26:404–12.
29. Hopper JL. Disease-specific prospective family study cohorts enriched for familial risk. *Epidemiol Perspect Innov* 2011;8:2.
30. Lange K. An approximate model of polygenic inheritance. *Genetics* 1997;147:1423–30.
31. Fernando RL, Stricker C, Elston RC. The finite polygenic mixed model: an alternative formulation for the mixed model of inheritance. *Theor Appl Genet* 1994;88:573–80.
32. Dowty JG, Win AK, Buchanan DD, Lindor NM, Macrae FA, Clendenning M, et al. Cancer risks for MLH1 and MSH2 mutation carriers. *Hum Mutat* 2013;34:490–7.
33. Baglietto L, Lindor NM, Dowty JG, White DM, Wagner A, Gomez Garcia EB, et al. Risks of Lynch syndrome cancers for MSH6 mutation carriers. *J Natl Cancer Inst* 2010;102:193–201.
34. Senter L, Clendenning M, Sotamaa K, Hampel H, Green J, Potter JD, et al. The clinical phenotype of Lynch syndrome due to germ-line PMS2 mutations. *Gastroenterology* 2008;135:419–28.
35. Win AK, Dowty JG, Cleary SP, Kim H, Buchanan DD, Young JP, et al. Risk of colorectal cancer for carriers of mutations in MUTYH, with and without a family history of cancer. *Gastroenterology* 2014;146:1208–11.e1–5.
36. Forman D, Bray F, Brewster DH, Gombe Mbalawa C, Kohler B, Piñeros M, et al. Cancer incidence in five continents. volume X. Lyon, France: International Agency for Research on Cancer, World Health Organization; 2014. Available from: <http://ci5.iarc.fr/CIS1-X/old/vol10/CISvol10.pdf>.
37. Akaike H. Information theory and an extension of the maximum likelihood principle. In: Parzen E, Tanabe K, Kitagawa G, editors. *Selected papers of Hirotugu Akaike*. New York: Springer New York; 1998. p. 199–213.
38. Bozdogan H. Model selection and Akaike's Information Criterion (AIC): the general theory and its analytical extensions. *Psychometrika* 1987;52:345–70.
39. Statistics Canada [database on internet]. Table: 13-10-0710-01. Deaths and mortality rates, by age group. CANSIM. doi: <https://doi.org/10.25318/1310071001-eng>. Available from: <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310071001>.
40. Australian Institute of Health and Welfare. *General record of incidence of mortality*. books 2013: All causes combined. 2015.
41. Miniño AM, Arias E, Kochanek KD, Murphy SL, Smith BL. *Deaths: final data for 2000*. National vital statistics reports; vol. 50 no. 15. Hyattsville, MD: National Center for Health Statistics; 2002. Available from: https://www.cdc.gov/nchs/data/nvsr/nvsr50/nvsr50_15.pdf.
42. Metzger R, Danenberg K, Leichman CG, Salonga D, Schwartz EL, Wadler S, et al. High basal level gene expression of thymidine phosphorylase (platelet-derived endothelial cell growth factor) in colorectal tumors is associated with nonresponse to 5-fluorouracil. *Clin Cancer Res* 1998;4:2371–6.
43. Freedman AN, Slattery ML, Ballard-Barbush R, Willis G, Cann BJ, Pee D, et al. Colorectal cancer risk prediction tool for white men and women without known susceptibility. *J Clin Oncol* 2009;27:686–93.
44. Tibshirani RJ, Efron B. An introduction to the bootstrap. *Monogr Stat Appl Probab* 1993;57:1–436.
45. Gray RJ. A class of K-sample tests for comparing the cumulative incidence of a competing risk. *Ann Stat* 1988;16:1141–54.
46. Balavarca Y, Weigl K, Thomsen H, Brenner H. Performance of individual and joint risk stratification by an environmental risk score and a genetic risk score in a colorectal cancer screening setting. *Int J Cancer* 2020;146:627–34.
47. Yasui Y, Newcomb PA, Trentham-Dietz A, Egan KM. Familial relative risk estimates for use in epidemiologic analyses. *Am J Epidemiol* 2006;164:697–705.
48. World Cancer Research Fund/American Institute for Cancer Research. *Food, nutrition, physical activity, and the prevention of cancer: a global perspective*. Washington DC: American Institute for Cancer Research; 2007.
49. Chlebowski RT, Wactawski-Wende J, Ritenbaugh C, Hubbell FA, Ascensao J, Rodabough RJ, et al. Estrogen plus progestin and colorectal cancer in postmenopausal women. *N Engl J Med* 2004;350:991–1004.
50. Peng L, Balavarca Y, Weigl K, Hoffmeister M, Brenner H. Head-to-head comparison of the performance of 17 risk models for predicting presence of advanced neoplasms in colorectal cancer screening. *Am J Gastroenterol* 2019;114:1520–30.
51. Jeon J, Du M, Schoen RE, Hoffmeister M, Newcomb PA, Berndt SI, et al. Determining risk of colorectal cancer and starting age of screening based on lifestyle, environmental, and genetic factors. *Gastroenterology* 2018;154:2152–64.e19.