

Risk Prediction Models for Colorectal Cancer: A Review

Aung Ko Win¹, Robert J. MacInnis^{1,2}, John L. Hopper¹, and Mark A. Jenkins¹

Abstract

Risk prediction models are important to identify individuals at high risk of developing the disease who can then be offered individually tailored clinical management, targeted screening and interventions to reduce the burden of disease. They are also useful for research purposes when attempting to identify new risk factors for the disease. In this article, we review the risk prediction models that have been developed for colorectal cancer and appraise their applicability, strengths, and weaknesses. We also discuss the factors to be considered for future development and improvement of models for colorectal cancer risk prediction. We conclude that there is no model that sufficiently covers the known risk factors for colorectal cancer that is suitable for assessment of people from across the full range of risk and that a new comprehensive model is needed. *Cancer Epidemiol Biomarkers Prev*; 21(3); 398–410. ©2011 AACR.

Introduction

Colorectal cancer (CRC) is one of the most frequently diagnosed cancers in the world with more than one million new cases diagnosed (9.8% of worldwide cancer diagnoses) and 600,000 deaths (8.1% of all worldwide cancer deaths) caused by the disease in 2008 (1). There is likely to be a large spectrum of risk across populations (2). While part of this is due to individual difference in exposure to environmental risk factors, in theory, there is also likely a large variation of risk due to underlying familial risk factors, which we refer to as the "familial risk profile" (3–7). Mathematical modeling of the relationship between family history as a risk factor and underlying familial risk profile suggests that (i) the risk of developing CRC varies approximately 20-fold between the people in the lowest quartile (average 1.25% lifetime risk of CRC) versus the highest quartile for familial risk profile (average 25% risk; ref. 4) and (ii) 90% of all CRCs occurs in people who are above the median familial risk profile (6, 8).

Risk factors for developing CRC

Many risk factors have been implicated in the development of CRC (9). Family history is a well-established risk factor (10). Studies have reported that individuals with one affected first-degree relative (parent, offspring, sibling) have, on average, a 2-fold increased risk of CRC

compared with those with no family history (10–12) and this association increases to 4-fold for individuals with 3 or more affected first-degree relatives (10). However, the above statistics are mere averages. Heterogeneity in both absolute and relative CRC cancer risk associated with family history depends on the age of the individual at risk, the age(s) at diagnosis of affected relative(s) (the earlier the age at diagnosis the more likely to be at risk), and the genetic relationship between them (the greater the number and/or the closer the relationship to, the more likely to be at risk; ref. 10).

Approximately 3% to 5% of CRC is now known to be due to mutations in the known high-risk cancer susceptibility genes (13). The strongest ones in terms of risk and mutation frequency comprise the DNA mismatch repair (MMR) genes that cause Lynch syndrome (14), the adenomatous polyposis coli gene that causes familial adenomatous polyposis (15), and the *MUTYH* gene that causes colorectal polyps and subsequently cancer (*MUTYH*-associated polyposis; ref. 16). Although rare (in aggregate, 1 in 500 to 1 in 10,000 of the population carry a mutation in these genes; refs. 17–19), these mutation carriers are at high risk of developing cancer, particularly CRC (20–28).

At most, only half of the familial risk of CRC can be explained by mutations in known high-risk genes (29). While much research capital has been spent on the search for other CRC "high-risk" susceptibility genes, none have been confirmed. On the other hand, genome-wide association studies have identified at least 15 common genetic susceptibility markers (single-nucleotide polymorphisms, SNP) that have been reliably shown to be associated with small increments in risk of developing CRC (ORs for homozygotes of the minor allele vs. noncarriers ranging from 0.80 to 1.70; refs. 30–38). The causes of these weak associations, let alone of the residual of the familial risk, are unknown, but they could include less common genetic variants and/or other risk factors shared by relatives.

Authors' Affiliation: ¹Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, The University of Melbourne, Parkville, Victoria, Australia; ²Cancer Epidemiology Centre, The Cancer Council Victoria, Carlton, Victoria, Australia

Corresponding Author: Mark A. Jenkins, Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, Melbourne School of Population Health, Level 3, 207 Bouverie Street, The University of Melbourne, Victoria 3010, Australia. Phone: 61-38344-0902; Fax: 61-39349-5815; E-mail: m.jenkins@unimelb.edu.au

doi: 10.1158/1055-9965.EPI-11-0771

©2011 American Association for Cancer Research.

The etiology of developing CRC is also attributable to nongenetic, personal, and environmental factors. Meta-analyses have reported positive associations with obesity [relative risk (RR) per 5 kg/m²: colon cancer, 1.24 (95% confidence interval, CI, 1.20–1.28) for men and 1.09 (95% CI, 1.05–1.13) for women; and rectal cancer, 1.09 (95% CI, 1.06–1.12) for men and 1.02 (95% CI, 1.00–1.05) for women; ref. 39], diabetes [RR, 1.29 (95% CI, 1.15–1.44) for men and 1.33 (95% CI, 1.23–1.44) for women; ref. 40], consumption of red meat [RR for 120 g/d, 1.28 (95% CI, 1.18–1.39); ref. 41] and processed meat [RR for 30 g/d, 1.09 (95% CI, 1.05–1.13); ref. 41], and negative associations with physical activity [RR of colon cancer for occupational activities vs. none, 0.79 (95% CI, 0.72–0.87) for men, and for recreational activities vs. none, 0.78 (95% CI, 0.68–0.91) for men and 0.71 (95% CI, 0.57–0.88) for women; ref. 42], postmenopausal hormone replacement therapy [RR for ever user vs. never user, 0.80 (95% CI, 0.74–0.86) for colon cancer and 0.81 (95% CI, 0.72–0.92) for rectal cancer; ref. 43], calcium intake [RR for the highest vs. the lowest quintile of intake, 0.86 (95% CI, 0.79–0.95) for dietary calcium and 0.78 (95% CI, 0.69–0.88) for total calcium; ref. 44], vitamin D intake [RR for the highest vs. lowest categories, 0.88 (95% CI, 0.80–0.96); ref. 45] and aspirin use [RR for ≥ 2 –3 times weekly for >1 year vs. never user, 0.78 (95% CI, 0.63–0.97); ref. 46].

Predicting Risk of CRC

Risk prediction models are important because they can be used (i) to reduce the burden of disease by enabling targeted preventive measures to those at highest risk; (ii) to identify those most likely to carry a genetic predisposition, who then can be triaged for genetic testing (47); and (iii) to increase the power of observation studies to identify new risk factors for the disease (provided that such risk prediction models do have high predictive power for identifying individuals at high risk for CRC). Risk prediction models are typically developed using data from observational study designs including case–control and cohort studies (48, 49).

Reducing the burden of disease

Health policy makers and clinicians rely on risk classifications to decide which individuals to screen for CRC. To date, these classification schemes are predominantly based only on age and/or a simple classification of family history. The advantage of using risk prediction models that incorporate multiple variables (including novel predictors such as genomic data) is that they generally have greater accuracy than clinical stage alone (50). CRC risk prediction models identify individuals most likely to benefit from CRC screening or other preventative interventions. The majority of CRCs derive from a polyp, a well-recognized premalignant form that may be present for many years before symptoms (such as rectal bleeding, change in bowel habit, or anemia) manifest. These polyps can be removed during colonoscopy, thereby reducing the

risk of CRC (51, 52). Furthermore, as CRCs are generally asymptomatic before the cancer has reached a relatively advanced stage, screening by fecal occult blood test and colonoscopy is an effective method to reduce metastasis and death by CRC by detecting the disease at an early stage (51, 53). Accurate CRC risk predictions could help with clinical decisions on the frequency and mode of CRC screening and preventive strategies for those who are at high risk while leaving those who are at low risk of disease unexposed to potentially invasive screening modalities.

Identification of carriers of a high-risk genetic susceptibility

A small proportion of the population have inherited a genetic mutation that puts them at a high risk of CRC (18, 19). The relevant genes identified to date include the MMR genes (*MLH1*, *MSH2*, *MSH6*, and *PMS2*), *APC*, *MUTYH*, *STK11 (LB1)*, *BMPR1A*, and *PTEN*. Carriers of a mutation in any one of these genes have a lifetime risk of CRC of 30% to 100% (compared with a 5% lifetime risk for the general population) and benefit greatly from CRC screening (54, 55). Likely carriers of a mutation in an MMR gene can be identified from among CRC cases by pathologic examination of their tumors (56), whereas likely carriers of a mutation in the *APC* gene and a proportion of carriers of 2 mutations in the *MUTYH* gene can be identified by colonoscopic examination of the bowel for multiple polyps (57). However, for CRC-unaffected individuals and those who are not screened by colonoscopy, the identification of likely carriers requires a predictive model. Once identified, those most likely to be carriers could be triaged to genetic testing (which would be prohibitively expensive if provided to the whole population given less than 1 in 500 of the population carry a mutation; ref. 19).

Identification of new risk factors for developing CRC

Traditionally, observational studies of CRC risk have either compared the frequency of environmental, lifestyle, and genetic risk factors of affected with unaffected (case–control design) or have compared CRC incidence for those exposed with that for those not exposed (cohort design). In the vast majority of these studies, participants have been sampled irrespective of their family history of CRC, and as a consequence, most of the population does not have a family history of the disease. Therefore, finding from such research might not be generalizable to people with a family history of the disease. Furthermore, as analyses are usually adjusted for a family history, the observed associations are merely the average association across the broad spectrum of familial risk profile (8). Consequently, little is known about the specific role of risk factors within any category of familial risk profile or whether the associations differ by familial risk profile.

Risk prediction models can be used as a tool to try to identify new risk factors for disease. One potential method would be to classify individuals into those who had high risk of disease and those with a low risk of disease and then to compare the cases and controls within these

low- and high-risk categories for a given exposure (provided it was not used to classify the risk). This method could have greater statistical power to identify additional risk factors for the disease and may provide evidence for gene-environment interactions (8).

Existing Risk Prediction Models for CRC

Several CRC risk prediction models have been developed on the basis of known genetic and environmental risk factors. The aim of this review is to summarize these models and the studies that have evaluated them in terms of their applicability, strengths, and weaknesses. Table 1 summarizes risk prediction models for CRC that incorporate personal, environmental, and lifestyle risk factors and/or cancer susceptibility genes. To focus our review, we excluded models that predict stage of CRC based on patient and tumor characteristics (e.g., Cai and colleagues; ref. 58), models that predict type of colorectal neoplasia (benign or malignant) based on clinical symptoms (e.g., Brazer and colleagues; ref. 59), models that predict CRC incidence based on polyp growth rates after first polypectomy (e.g., Wilson and Lightwood; ref. 60), models that predict symptomatic CRC based on bowel symptoms of a patient assessed by a general practitioner (e.g., Selvachandran and colleagues; ref. 61), and models that only predict risk of metachronous CRC (i.e., subsequent primary cancer) based on characteristics of first cancer and gene signature (e.g., Peng and colleagues; ref. 62).

"Non-genetic" models

Most of the previously developed models that predict CRC risk use scoring systems based on regression models incorporating family history, lifestyle, and environmental risk factors. Such models are straightforward to implement and nongenetic risk factors can readily be incorporated. However, it is difficult to incorporate detailed information on family history including number of relatives, ages of unaffected relatives, ages of diagnoses, beyond first-generation relatives, and risk factors correlated between family members (both genetic and nongenetic). Consequently, these "nongenetic" models may not give precise or accurate estimates of future CRC risk for important groups. For example, while the models that take account family history of colon cancer (63) or CRC (64, 65) as a binary variable (yes/no) provide an average risk for groups of individuals at a low risk, they are unable to provide accurate risk estimates for individuals at high-risk due to having a strong family history or carrying a known genetic mutation.

Another important issue with respect to history of colonoscopy is the occurrence of screening or diagnostic colonoscopies. A negative colonoscopy indicates a reduced risk of disease (compared with someone who has had no colonoscopy; ref. 66) and a positive colonoscopy indicates an increased risk of disease (evidence of susceptibility to further adenomas; ref. 67) as well as a potential decreased risk of disease if it was followed by

polypectomy to remove pre-cancerous adenoma (68). Currently, it is not clear how to incorporate such information into a risk prediction model, and this will be important as colonoscopies are being increasingly used as a diagnostic or screening test.

Genetic models

We are aware of only one model that bases predictions on inherited genetic risk factors for the disease. MMRpro (previously known as CRCAPRO; ref. 69) was developed by Chen and colleagues (70) and it assumes that all familial aggregation is due to dominantly inherited, highly penetrant mutations in *MLH1*, *MSH2*, or *MSH6*. It predicts the probability of being a mutation carrier, by gene, as well as age-specific CRC (and endometrial cancer) risks for unaffected individuals. This model assumes a mutation carrier frequency of 0.0009 (~1 in 1,100) for *MLH1*, 0.0010 (~1 in 1,000) for *MSH2*, and 0.00036 (~1 in 2,800) for *MSH6* with penetrance functions based on published estimates (70). There are several limitations of this model: (i) it uses family history of CRC only up to second-degree relatives (Table 1); (ii) it does not incorporate the MMR gene *PMS2*, which accounts for 15% of MMR gene mutations (71), although this probably has little impact given risk estimates are robust to mutation frequencies in this range; (iii) it does not include any environmental risk factors; and (iv) it does not predict second primary cancer risk for affected individuals.

None of the models described above (summarized in Table 1) adequately account for any complexities of the genetic susceptibility of CRC. For example, high-risk mutations in the known CRC susceptibility genes only account for half of the familial risk at most (29). Genome-wide association studies have identified SNPs located in 15 loci that individually are associated with small increments in risk of CRC (37, 72). Modeling suggests that should the residual familial risk be explained by similarly weak genetic effects, then hundreds of genes may exist that alter CRC risk (73). While cancer prediction models have been developed to account for a polygenic component representing the effects of a large number of genetic variants each of small effect on risk for breast (74, 75) and prostate cancers (76), no comparable tool has been developed for CRC.

Evaluation of Risk Prediction Models

Before a risk prediction model can be recommended as a useful tool for individualized decision making in a clinical setting, it needs to be validated in an independent population than that was used to create the model (77). The most important characteristics used to evaluate the performance of a risk prediction model are as follows:

- (i) *Calibration* (or *reliability*) assesses the ability of a model to predict the number of events (CRCs in this setting) it purports to predict. This is commonly described by using the goodness-of-fit or χ^2 statistic to compare the

Table 1. Summary of the previously developed CRC risk prediction models

Model	Study design and sample	Methods	Factors included					Residual risk factors ^a	Applicability	Strengths	Limitations
			Family history	Environmental factors	High-risk genetic mutations	Environmental factors	High-risk genetic mutations				
Colditz and colleagues (63) Harvard Cancer Risk Index	Estimated parameters from published data and expert opinion	Scoring for each risk factor based on strengths of associations from the previous logistic regression analyses	FDR with colon cancer (yes/no)	BMI, screening (FOBT and sigmoidoscopy), aspirin, inflammatory bowel disease, folate, vegetables, alcohol, height, physical activity, estrogen replacement, OC, red meat, fruits, fiber, saturated fat, cigarette smoking	None	None	None	Predicts 10-y risk of colon cancer	Ease of use	May not be applicable to rectal cancer risk prediction. Does not consider family history in relatives beyond first degree and is not applicable for people with a high-risk genetic mutation.	
Imperiale and colleagues (64)	A cross-sectional study of 1,994 asymptomatic individuals aged ≥ 50 years identified between 1995 and 2001	Scoring for each risk factor (method for score undefined)	None	Age, sex, most advanced distal nonmalignant neoplasm (no polyps; hyperplasia; tubular adenoma <1 cm; advanced lesion—tubular adenoma >1 cm, any polyp with villous histology or severe dysplasia or cancer)	None	None	Determines the need for screening colonoscopy based on the findings of a flexible sigmoidoscopy. Predicts risk of proximal colon cancer.	Improves efficiency of CRC screening	Limited to individuals having a sigmoidoscopy. May not be applicable to people aged younger than 50y. Is not applicable for people with strong family history or with a high-risk genetic mutation.		
Driver and colleagues (80)	A prospective cohort of 21,581 men aged 40 to 84 years (Physician's Health Study); followed-up from 1982 to 2004; 485 incident CRC cases.	Logistic regression	None	Age, smoking, alcohol, BMI [diabetes, physical activity, vegetables, cold cereal multivitamins, vitamin C, and vitamin E were considered but not included in the model]	None	None	Provides CRC risk score. Predicts 20-y risk of CRC for men	Ease of use. Based on large sample	Limited to males. Is not applicable to people with a strong family history or with a high-risk genetic mutation.		

(Continued on the following page)

Table 1. Summary of the previously developed CRC risk prediction models (Cont'd)

Model	Study design and sample	Methods	Factors included					Residual risk factors ^a	Strengths	Limitations
			Family history	Environmental factors	High-risk genetic mutations	Applicability	Strengths			
Freedman and colleagues (64)	2,263 cases and 2,833 controls of non-Hispanic white men and women aged ≥ 50 years identified between 1991–1994 (colon) and 1997–2001 (rectal)	Logistic regression	FDR with CRC (yes/no), number of FDR with CRC (0, 1, ≥ 2)	Age, sex, sigmoidoscopy and colonoscopy, current leisure time activity, aspirin and NSAIDs, cigarette smoking, vegetables, BMI, and hormone replacement	None	None	Predicts 5-, 10-, and 20-y, and lifetime risk of developing CRC for men and women older than 50 y	User friendly web version available (108) Based on large sample	May not be applicable to people younger than 50 y. Does not consider family history in relatives beyond first degree. Is not applicable to people with a strong family history or with a high-risk genetic mutation	
Wei and colleagues (65)	A prospective cohort of 83,767 women aged 30 to 54 years (Nurses' Health Study); follow-up from 1976 to 2004; 701 incident colon cancer cases.	Nonlinear Poisson regression	FDR with CRC (yes/no)	Age, sigmoidoscopy and colonoscopy, physical activity, aspirin, cigarette smoking, processed meat or red meat, folate, height, BMI, and hormone replacement	None	None	Predicts cumulative risk of colon cancer for women aged 30 to 70 y	Ease of use. Based on large sample	May not be applicable to men or to any women aged older than 70 y. Does not consider family history in relatives beyond first degree. Is not applicable to people with a strong family history or with a high-risk genetic mutation. May not be applicable for rectal cancer risk	
Ma and colleagues (63)	A prospective cohort of 28,115 men aged 40 to 69 years (Japan Public Health Center-based study—Cohort II); followed-up from 1993 to 2005 (mean, 11.0 y); 543 incident CRC cases.	Cox proportional hazards model	None [FDR with CRC (yes/no) was considered, but not included in the model]	Age, BMI, physical activity, smoking, alcohol [diabetes was considered but not included in the model]	None	None	Provides CRC risk score and predicts 10-y risk of CRC for Japanese men.	Ease of use	May not be applicable to non-Japanese or to Japanese females. Is not applicable to people with strong family history or with a high-risk genetic mutation	

(Continued on the following page)

Table 1. Summary of the previously developed CRC risk prediction models (Cont'd)

Model	Study design and sample	Methods	Factors included				Residual risk factors ^a	Applicability	Strengths	Limitations
			Family history	Environmental factors	High-risk genetic mutations	Environmental factors				
Chen and colleagues (70) <i>MMRpro</i>	Estimated parameters from published data	Bayesian/segregation analysis	FDR and SDR: specific relationship to the proband, history of CRC/EC (yes, no), age at diagnosis, MSI status, and <i>MLH1</i> , <i>MSH2</i> , <i>MSH6</i> mutation status	Age, race/ethnicity,	<i>MLH1</i> , <i>MSH2</i> , <i>MSH6</i>	None	Predicts probability of carrying <i>MLH1</i> , <i>MSH2</i> , and <i>MSH6</i> mutations Predicts 5-y and lifetime risk of CRC and EC	Included specific family history to second-degree. Uses stand-alone software package that is freely available	Does not consider family history in relatives beyond second degree. May not be applicable to <i>PMS2</i> mutation carriers. Does not estimate risk of second primary (metachronous) CRC. Is not applicable to <i>MUTYH</i> mutation carriers.	
Cleveland Clinic Tool (109)	Unreported	Unknown	FDR and SDR: specific relationship to the proband, history of CRC and polyps (yes, no), age at diagnosis of CRC (<50, 50-60, ≥60), age at diagnosis of polyps (<60, ≥60)	Age (<50, ≥50), sex, ethnicity, weight, height, CRC screening (colonoscopy, sigmoidoscopy, FOBT), fruit and vegetables consumption, smoking, exercise, person history of CRC and polyps	None	None	Provides CRC risk score (average/medium/high)	User friendly web version available (109)	Not possible to assess as methods used for development of this tool have not been published. Does not predict cumulative risk over a specified period	

Abbreviations: BMI, body mass index; EC, endometrial cancer; FDR, first-degree relative; FOBT, fecal occult blood test; MSI, microsatellite instability of tumor; NSAID, nonsteroidal anti-inflammatory drugs; OC, oral contraceptive; SDR, second-degree relative.
^aFamilial aggregation that is not explained by known risk factors (including genetic and environmental risk factors shared by family members).

Table 2. Summary of studies that evaluated CRC risk prediction models

Study	Model evaluated	Sample	Features evaluated	Key findings
Kim and colleagues (85)	Harvard Caner Risk Index (63)	A prospective cohort of 38,953 men aged 40 to 70 years (Health Professionals Follow-up Study); follow-up from 1986 to 1996; 230 incident CRC cases. A prospective cohort of 52,668 women aged 40 to 70 years (Nurses' Health Study); follow-up from 1984 to 1994; 244 incident CRC cases.	Calibration, discrimination, and utility	Overestimated the number of CRC for men within "much below average risk" and "very much below average risk" risk categories. Well calibrated for women. Modest discrimination: $c = 0.71$ (95% CI, 0.68–0.74) for men and 0.67 (95%CI, 0.64–0.70) for women.
Emmons and colleagues (86)	Harvard Caner Risk Index (63)	In-depth cognitive interviews with 9 individuals from the general population; and 9 focus groups (6 females and 3 males) aged ≥ 40 years	Meaning of risk, perceptions about cancer, and interpretation of the results	66% extremely/very satisfied, 32% somewhat satisfied with the model format. 86% extremely/very satisfied, 13% somewhat satisfied with the information provided in the model. 3% not at all satisfied with the model. Some dissatisfied because exposures that they believed to be important were not included (e.g., poverty, toxic waste, air pollution) Difficult for participants in completing the HCRI in its paper-and-pencil form.
Emmons and colleagues (87)	Harvard Colorectal Cancer Risk Assessment and Communication Tool for Research (HCCRACT-R; ref. 63)	A randomized control trial on 159 men and 194 women aged 40 to 70 years without previous personal history of cancer. Intervention groups: those receiving (i) presentation of both absolute and relative risk (ii) presentation of absolute risk only Control group: those without receiving any personal risk information	Accuracy of risk perception, and, level of worry and satisfaction	Significant changes in risk perception accuracy for both relative risk ($P = 0.01$) and absolute risk ($P = 0.001$) across intervention groups. Of those with inaccurate absolute risk perception at baseline, 54% of the participants in the group who received presentation of both absolute and relative risk, and 64% of those in the group who received presentation of absolute risk only, had correct absolute risk perception at post-test, compared with only 12% of the control group. 13% less worried, 17% more worried about getting CRC after completing the HCCRACT-R.

(Continued on the following page)

Table 2. Summary of studies that evaluated CRC risk prediction models (Cont'd)

Study	Model evaluated	Sample	Features evaluated	Key findings
Imperiale and colleagues (84)	Imperiale (84)	A cross-sectional study on 1,031 asymptomatic individuals aged ≥ 50 years undergoing first-time screening colonoscopy between 1999 and 2001	Discrimination	Modest discrimination: $c = 0.74$ (SD = 0.06)
Park and colleagues (88)	Freedman (64)	A prospective cohort of 155,345 men and 108,057 women aged 50 to 71 years (American Association of Retired Persons—diet and health study); follow-up from 1995 to 2003 (mean, 6.9 y); 2,092 male and 965 female incident CRC cases.	Calibration and discrimination	Well calibrated for men (E/O = 0.99; 95% CI, 0.95–1.04) and for women (E/O = 1.05; 95% CI, 0.98–1.11) overall. Overestimated risk for men with one affected relative (E/O = 1.35; 95% CI, 1.17–1.55); women with one affected relative (E/O = 1.20; 95% CI, 1.00–1.45); men with 2 affected relatives (E/O = 1.48; 95% CI, 1.00–2.19); and men who had a history of screening and polyps (E/O = 1.42; 95% CI, 1.24–1.63). Underestimated risk for men who had a history of screening with no polyps (E/O = 0.67; 95% CI, 0.62–0.72). Modest discriminatory accuracy: $c = 0.61$ (95% CI, 0.60–0.62) for men and 0.61 (95% CI, 0.59–0.62) for women
Ma and colleagues (83)	Ma (83)	A prospective cohort of 18,256 men aged 40 to 59 years (Japan Public Health Center-based study—Cohort I); follow-up from 1990 to 2005 (mean, 10.1 y); 389 incident CRC cases.	Calibration and discrimination	Underestimation for colon cancer: O/E = 1.19 (95% CI, 1.03–1.37). Well calibrated for rectal cancer (O/E = 0.94; 95% CI, 0.78–1.12) and for CRC overall (O/E = 1.09; 95% CI, 0.98–1.23) Modest discriminatory accuracy: $c = 0.64$ (95% CI, 0.61–0.67) for CRC, 0.66 (95% CI, 0.62–0.70) for colon cancer, 0.62 (95% CI, 0.57–0.66) for rectal cancer

Abbreviations: c, concordance statistic; E, expected; O, observed.

expected number of events with the observed number of events (78).

- (ii) *Discrimination* (or *precision*) measures the ability of a model to distinguish individuals more likely to develop disease from those less likely to develop disease, by using the concordance statistic (*c*-statistic) that corresponds to the area under a receiver operating characteristic curve (AUC; ref. 78) or the net reclassification index (NRI), which is the probability

that the enhanced model reclassifies a person correctly (i.e., increases the estimate of risk of a person who develops CRC or decreases the estimate of risk of a person who does not) minus the probability of an incorrect reclassification (i.e., decreases the estimate of risk of a person who develops CRC or increases the estimate of risk of one who does not; ref. 79).

- (iii) *Accuracy* evaluates how well a model categorizes specific individuals into those more and less likely to

develop disease and the usefulness on an individual basis in predicting disease risk. Sensitivity and specificity as well as positive and negative predictive values are important indices for this test.

- (iv) *Utility*, which is the ability of the model to be completed by those it is designed for, such as clinicians, patients, general population, and health policy makers. This is commonly evaluated by results from surveys or interviews of users.

Table 2 summarizes the studies that have evaluated the models for CRC risk prediction. To our knowledge, the models of Driver and colleagues (80) and Wei and colleagues (65) have not been validated. For MMRpro (70), some studies (e.g., refs. 81, 82) have evaluated the probability of a mismatch repair gene mutation status, but not CRC risk. The other 4 models (63, 64, 83, 84) were validated using an independent data set from a separate study or a subset of the same data set used to generate a model. There have been no publications comparing these models with each other.

Harvard Cancer Risk Index

Three studies (85–87) have assessed this model including 2 qualitative studies (86, 87). While this model was well calibrated for women, it overestimated the number of CRC for men at low risk (85). Utility was evaluated by assessing the lay understanding of risk, perception of risk and interpretation of the results, and was determined to be well received by users (86). A computer-based tool that provides an estimate of personal absolute and relative risk for CRC (the Harvard Colorectal Cancer Risk Assessment and Communication Tool for Research) was evaluated qualitatively by Emmons and colleagues (87) for accuracy of risk perception and level of worry and satisfaction by users. They concluded that the tool was useful for correcting misperceptions about personal risk. Of those with inaccurate risk perception at baseline, more than half of the participants in intervention groups had corrected risk perceptions at posttest, compared with only 12% in the control group (Table 2).

Imperiale's model

Imperiale and colleagues (84) validated their own model using an independent data set, from the same source used to develop the model; 1,031 men and women without bowel symptoms who underwent a screening colonoscopy. They observed that the respective actual risk of advanced lesion in the proximal colon (tubular adenomas larger than 1 cm, any polyp with villous histology or severe dysplasia, or cancer) was similar to the predicted low-, intermediate- and high-risk estimated by the model, albeit with modest discrimination indicated by the c-statistic of 0.74 (84). However, the validity of the model for different populations is unknown, given that both the model and validation data sets were from the same population.

Freeman's model

Park and colleagues (88) evaluated Freeman's model (64) using prospective data from a cohort of ~260,000 individuals over a mean follow-up of 7 years. They observed that Freeman's model was well calibrated overall for both men and women and in most categories of risk factors. However, the model overestimated risk for individuals with a family history of CRC, with the expected number of CRCs being 35% (95% CI, 17%–55%) higher than the observed number for men who had one affected relative with CRC, and 42% (24%–63%) higher for men who had a history of screening and polyps. The model underestimated the risk for men who had a history of screening but no polyps by 33% (95% CI, 28%–38%; Table 2). The model had a modest discrimination indicated by the c-statistic of 0.61 for men and women.

Ma's model

Ma and colleagues (83) validated their own model using the Japan Public Health Center-based study. They found that their model underestimated the number of colon cancers by 19% (95% CI, 3%–37%), but there was good agreement for rectal cancer and CRC overall. This model underestimated CRC cases with 4 categories of environmental exposures (age, body mass index, alcohol consumption, and smoking).

Future Perspective

The models that do exist for CRC risk are limited in terms of incorporated risk factors and their usefulness in terms of validity. More complex modeling is required, based on known risk factors for disease AND hypothesized but residual risk factors. In this section, we discuss the issues to consider for developing a comprehensive model to predict CRC risk.

Mutation-specific risks

Risk of CRC varies greatly depending on the presence or absence of a germ line mutation in a cancer risk susceptibility gene. For example, according to the studies that appropriately conditioned on ascertainment of mutation carrying families, the carriers of mutations in MMR genes are estimated to have at least 10 times higher cumulative risk (penetrance) to age 70 years than for the general population: *MLH1* and *MSLH2*, 56% (95% CI, 37%–75%) for men and 48% (95% CI, 26%–65%) for women (21); *MSH6*, 22% (95% CI, 14%–32%) for men and 10% (95% CI, 5%–17%) for women (20); and *PMS2*, 20% (95% CI, 11%–34%) for men and 15% (95% CI, 8%–26%) for women (22). Therefore, it is important that future models, particularly those to be applied to populations with a proportion of people who carry these mutations, should have the ability to derive genotype-specific cancer risks. Only the MMRpro model (70) incorporates such risks for *MLH1*, *MSH2*, and *MSH6* mutations.

Environmental factors

The role of physical characteristics and environmental exposures on CRC risk could depend on the presence or absence of carrying an inherited genetic mutation (89). Previous models have not accounted for gene–environment interactions or gene–gene interactions. Incorporation of these interactions is challenging. Studies have observed that the strengths of associations between environmental risk factors and CRC for those with a family history differs compared with those randomly selected from the population (90–92), which is consistent with the existence of gene–environment interactions. However to date, there have been few studies directly comparing strengths of associations for carriers with that for non-carriers. We have shown that body mass index in early adulthood is similarly associated with CRC risk for both carriers and noncarriers (93). Studies of alcohol consumption (94), fruit consumption (95), dietary fiber intake (95), smoking (95, 96) have not directly compared strengths of associations for carriers with noncarriers.

Need for ethnicity-specific risk models

Most existing CRC risk prediction models were developed using data from predominantly Caucasian populations and therefore they may not be applicable to other racial/ethnic populations. Further studies are needed for such populations. There is a well-recognized high degree of heterogeneity by nationality in CRC incidence with an up to 10-fold difference internationally (1, 2). Although some genetic differences may exist for CRC risk, studies have suggested that much of this difference may be due to differences in environmental risk factors as the incidence rate of CRC in migrants approach that of the host country within one or two generations (97).

Subtypes of CRC

Risk prediction models need to incorporate different subtypes of CRC (proximal colon, distal colon, and rectal cancer) as distinct disease endpoints as it is possible that the genetic and environmental risk factors may differ by anatomic location. For instance, beer consumption is associated with rectal cancer in men but not in women, whereas there is little evidence of an association of any dimension of alcohol consumption being associated with the risk of colon cancer (98).

Incorporation of metachronous CRC risk prediction

The incidence of metachronous CRC, that is, a subsequent primary cancer for people with a previous diagnosis of CRC, is estimated to be 8% over 5 years (99, 100), which is higher than the incidence of CRC for people who

have never been diagnosed with CRC (0.6% at age 50 years and 5.1% at age 70 years over 5 years). Future risk prediction models will need to be extended to include estimates of cancer risk subsequent to the first diagnosis if they are to be useful to those with a previous diagnosis of CRC.

Extracolonic cancers

Inherited CRC predisposition syndromes are seldom confined to the colorectum. Increased risks of various extracolonic cancers have been reported: cancers of the uterus, stomach, ovary, ureter, renal pelvis, brain, small bowel, and hepatobiliary tract in Lynch syndrome (101); cancers of the brain, thyroid, and liver in familial adenomatous polyposis (102); and cancers of the uterus and stomach in monoallelic *MUTYH* mutation carriers (103), and cancers of the duodenum, bladder, skin, and ovary in biallelic *MUTYH* mutation carriers (104). Even within families not known to be segregating a high-risk mutation, CRC risk is higher for those with a family history of extracolonic cancers (105–107). The effects of these other cancers in relatives on CRC risk should therefore be incorporated into risk models.

Conclusions

To determine individual risk of CRC precisely, it is important to develop a genetically based prediction model that incorporates the multigeneration family history and environmental risk factors, and allowing for the simultaneous effects of different MMR genes and other high cancer risk susceptibility genes including *MUTYH*, and the effect of low-penetrance genes and residual genetic or other familial risk factors. Only such a model can provide precise estimates of future cancer risks, and predict high-risk genetic mutation status for each individual across the full CRC risk spectrum. Validated risk prediction models should be easily assessable and user friendly for clinicians and genetic counselors as well as the general public by implementing them as a web-based application.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interests were disclosed.

Grant Support

A.K. Win was supported by a grant from the Cancer Council Victoria and the Picchi Brothers Foundation, Australia. R.J. MacInnis is a Sidney Sax Post Doctoral Research Fellow and J.L. Hopper is an Australia Fellow of the National Health and Medical Research Council (NHMRC), Australia.

Received August 11, 2011; revised December 5, 2011; accepted December 6, 2011; published OnlineFirst December 14, 2011.

References

1. Ferlay J, Shin H, Bray F, Forman D, Mathers C, Parkin D. GLOBOCAN 2008, cancer incidence and mortality worldwide. Lyon, France: International Agency for Research on Cancer; 2010.
2. Jemal A, Center MM, DeSantis C, Ward EM. Global patterns of cancer incidence and mortality rates and trends. *Cancer Epidemiol Biomarkers Prev* 2010;19:1893–907.

3. Aalen OO. Modelling the influence of risk factors on familial aggregation of disease. *Biometrics* 1991;47:933–45.
4. Hopper JL, Carlin JB. Familial aggregation of a disease consequent upon correlation between relatives in a risk factor measured on a continuous scale. *Am J Epidemiol* 1992;136:1138–47.
5. Khoury MJ, Beaty TH, Liang KY. Can familial aggregation of disease be explained by familial aggregation of environmental risk factors? *Am J Epidemiol* 1988;127:674–83.
6. Pharoah PD, Antoniou A, Bobrow M, Zimmern RL, Easton DF, Ponder BA. Polygenic susceptibility to breast cancer and implications for prevention. *Nat Genet* 2002;31:33–6.
7. Peto J, Mack TM. High constant incidence in twins and other relatives of women with breast cancer. *Nat Genet* 2000;26:411–4.
8. Hopper JL. Disease-specific prospective family study cohorts enriched for familial risk. *Epidemiol Perspect Innov* 2011;8:2.
9. Potter JD. Risk factors for colon neoplasia—epidemiology and biology. *Eur J Cancer* 1995;31A:1033–8.
10. 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.
11. Johns LE, Houlston RS. A systematic review and meta-analysis of familial colorectal cancer risk. *Am J Gastroenterol* 2001;96:2992–3003.
12. Baglietto L, Jenkins MA, Severi G, Giles GG, Bishop DT, Boyle P, et al. Measures of familial aggregation depend on definition of family history: meta-analysis for colorectal cancer. *J Clin Epidemiol* 2006;59:114–24.
13. Rustgi AK. The genetics of hereditary colon cancer. *Genes Dev* 2007;21:2525–38.
14. Vasen HFA. Clinical diagnosis and management of hereditary colorectal cancer syndromes. *J Clin Oncol* 2000;18:81S–92S.
15. Fearhead NS, Britton MP, Bodmer WF. The ABC of APC. *Hum Mol Genet* 2001;10:721–33.
16. Al-Tassan N, Chmiel NH, Maynard J, Fleming N, Livingston AL, Williams GT, et al. Inherited variants of MYH associated with somatic G:C→T:A mutations in colorectal tumors. *Nat Genet* 2002;30:227.
17. Win AK, Hopper JL, Jenkins MA. Association between monoallelic MUTYH mutation and colorectal cancer risk: a meta-regression analysis. *Fam Cancer* 2011;10:1–9.
18. Dunlop MG, Farrington SM, Nicholl I, Aaltonen L, Petersen G, Porteous M, et al. Population carrier frequency of hMSH2 and hMLH1 mutations. *Br J Cancer* 2000;83:1643–5.
19. Hampel H, de la Chapelle A. The search for unaffected individuals with Lynch syndrome: do the ends justify the means? *Cancer Prev Res* 2011;4:1–5.
20. 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.
21. Jenkins MA, Baglietto L, Dowty JG, Vliet CMV, Smith L, Mead LJ, et al. Cancer risks for mismatch repair gene mutation carriers: a population-based early onset case-family study. *Clin Gastroenterol Hepatol* 2006;4:489–98.
22. 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.e1.
23. Spirio L, Otterud B, Stauffer D, Lynch H, Lynch P, Watson P, et al. Linkage of a variant or attenuated form of adenomatous polyposis coli to the adenomatous polyposis coli (APC) locus. *Am J Hum Genet* 1992;51:92–100.
24. Petersen GM, Slack J, Nakamura Y. Screening guidelines and pre-morbid diagnosis of familial adenomatous polyposis using linkage. *Gastroenterology* 1991;100:1658–64.
25. Cleary SP, Cotterchio M, Jenkins MA, Kim H, Bristow R, Green R, et al. Germline MutY human homologue mutations and colorectal cancer: a multisite case-control study. *Gastroenterology* 2009;136:1251–60.
26. Lubbe SJ, Di Bernardo MC, Chandler IP, Houlston RS. Clinical implications of the colorectal cancer risk associated with MUTYH mutation. *J Clin Oncol* 2009;27:3975–80.
27. Jenkins MA, Croitoru ME, Monga N, Cleary SP, Cotterchio M, Hopper JL, et al. Risk of colorectal cancer in monoallelic and biallelic carriers of MYH mutations: a population-based case-family study. *Cancer Epidemiol Biomarkers Prev* 2006;15:312–4.
28. Bonadona V, Bonaiti B, Olschwang S, Grandjouan S, Huiart L, Longy M, et al. Cancer risks associated with germline mutations in MLH1, MSH2, and MSH6 genes in Lynch syndrome. *JAMA* 2011;305:2304–10.
29. Aaltonen L, Johns L, Järvinen H, Mecklin J-P, Houlston R. Explaining the familial colorectal cancer risk associated with mismatch repair (MMR)-deficient and MMR-stable tumors. *Clin Cancer Res* 2007;13:356–61.
30. Tomlinson I, Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, Spain S, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet* 2007;39:984–8.
31. Haiman CA, Le Marchand L, Yamamoto J, Stram DO, Sheng X, Kolonel LN, et al. A common genetic risk factor for colorectal and prostate cancer. *Nat Genet* 2007;39:954–6.
32. Tenesa A, Farrington SM, Prendergast JGD, Porteous ME, Walker M, Haq N, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat Genet* 2008;40:631–7.
33. Zanke BW, Greenwood CMT, Rangrej J, Kustra R, Tenesa A, Farrington SM, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet* 2007;39:989–94.
34. Houlston RS, Webb E, Broderick P, Pittman AM, Di Bernardo MC, Lubbe S, et al. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat Genet* 2008;40:1426–35.
35. Jaeger E, Webb E, Howarth K, Carvajal-Carmona L, Rowan A, Broderick P, et al. Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. *Nat Genet* 2008;40:26–8.
36. Broderick P, Carvajal-Carmona L, Pittman AM, Webb E, Howarth K, Rowan A, et al. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nat Genet* 2007;39:1315–7.
37. Houlston RS, Cheadle J, Dobbins SE, Tenesa A, Jones AM, Howarth K, et al. Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. *Nat Genet* 2010;42:973–7.
38. Tomlinson IPM, Webb E, Carvajal-Carmona L, Broderick P, Howarth K, Pittman AM, et al. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat Genet* 2008;40:623–30.
39. Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet* 2008;371:569–78.
40. Larsson SC, Orsini N, Wolk A. Diabetes mellitus and risk of colorectal cancer: a meta-analysis. *J Natl Cancer Inst* 2005;97:1679–87.
41. Larsson SC, Wolk A. Meat consumption and risk of colorectal cancer: a meta-analysis of prospective studies. *Int J Cancer* 2006;119:2657–64.
42. Samad AK, Taylor RS, Marshall T, Chapman MA. A meta-analysis of the association of physical activity with reduced risk of colorectal cancer. *Colorectal Dis* 2005;7:204–13.
43. Grodstein F, Newcomb PA, Stampfer MJ. Postmenopausal hormone therapy and the risk of colorectal cancer: a review and meta-analysis. *Am J Med* 1999;106:574–82.
44. Cho E, Smith-Warner SA, Spiegelman D, Beeson WL, van den Brandt PA, Colditz GA, et al. Dairy foods, calcium, and colorectal cancer: a pooled analysis of 10 cohort studies. *J Natl Cancer Inst* 2004;96:1015–22.
45. Ma Y, Zhang P, Wang F, Yang J, Liu Z, Qin H. Association between vitamin D and risk of colorectal cancer: a systematic review of prospective studies. *J Clin Oncol* 2011;29:3775–82.
46. Dube C, Rostom A, Lewin G, Tsertsivadze A, Barrowman N, Code C, et al. The use of aspirin for primary prevention of colorectal cancer: a

- systematic review prepared for the U.S. Preventive Services Task Force. *Ann Intern Med* 2007;146:365–75.
47. Weitzel JN, Blazer KR, Macdonald DJ, Culver JO, Offit K. Genetics, genomics, and cancer risk assessment: state of the art and future directions in the era of personalized medicine. *CA Cancer J Clin* 2011 Aug 19. [Epub ahead of print].
 48. Harrell FE. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. New York: Springer-Verlag; 2001.
 49. Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating. New York: Springer; 2009.
 50. Vickers AJ. Prediction models in cancer care. *CA Cancer J Clin* 2011 Jun 23. [Epub ahead of print].
 51. Levin B, Lieberman DA, McFarland B, Smith RA, Brooks D, Andrews KS, et al. Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *CA Cancer J Clin* 2008;58:130–60.
 52. Smith RA, Cokkinides V, Brooks D, Saslow D, Shah M, Brawley OW. Cancer screening in the United States, 2011: a review of current American Cancer Society guidelines and issues in cancer screening. *CA Cancer J Clin* 2011;61:8–30.
 53. Winawer SJ, Zauber AG, Fletcher RH, Stillman JS, O'Brien MJ, Levin B, et al. Guidelines for colonoscopy surveillance after polypectomy: a consensus update by the US Multi-Society Task Force on colorectal cancer and the American Cancer Society. *CA Cancer J Clin* 2006;56:143–59.
 54. Jarvinen HJ, Aarnio M, Mustonen H, Aktan-Collan K, Aaltonen LA, Peltomaki P, et al. Controlled 15-year trial on screening for colorectal cancer in families with hereditary nonpolyposis colorectal cancer. *Gastroenterology* 2000;118:829–34.
 55. de Jong AE, Hendriks YMC, Kleibeuker JH, de Boer SY, Cats A, Griffioen G, et al. Decrease in mortality in Lynch syndrome families because of surveillance. *Gastroenterology* 2006;130:665–71.
 56. Southey MC, Jenkins MA, Mead L, Whitty J, Trivett M, Tesoriero AA, et al. Use of molecular tumor characteristics to prioritize mismatch repair gene testing in early-onset colorectal cancer. *J Clin Oncol* 2005;23:6524–32.
 57. Buchanan D, Young J. A perspective on bi-allelic MUTYH mutations in patients with hyperplastic polyposis syndrome. *Gastroenterology* 2009;136:2407–8.
 58. Cai G, Xu Y, Zhu X, Peng J, Li Z, Xiao C, et al. A simple scoring system based on clinical features to predict locally advanced rectal cancers. *J Gastrointest Surg* 2009;13:1299–305.
 59. Brazer SR, Pancotto FS, Long TT III, Harrell FE Jr, Lee KL, Tyor MP, et al. Using ordinal logistic regression to estimate the likelihood of colorectal neoplasia. *J Clin Epidemiol* 1991;44:1263–70.
 60. Wilson L, Lightwood J. Model of estimated rates of colorectal cancer from polyp growth by year of surveillance. *J Med Screen* 2001;8:187.
 61. Selvachandran SN, Hodder RJ, Ballal MS, Jones P, Cade D. Prediction of colorectal cancer by a patient consultation questionnaire and scoring system: a prospective study. *Lancet* 2002;360:278–83.
 62. Peng J, Wang Z, Chen W, Ding Y, Wang H, Huang H, et al. Integration of genetic signature and TNM staging system for predicting the relapse of locally advanced colorectal cancer. *Int J Colorectal Dis* 2010;25:1277–85.
 63. Colditz G, Atwood K, Emmons K, Monson R, Willett W, Trichopoulos D, et al. Harvard report on cancer prevention volume 4: Harvard Cancer Risk Index. *Cancer Causes Control* 2000;11:477–88.
 64. Freedman AN, Slattery ML, Ballard-Barbash 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.
 65. Wei EK, Colditz GA, Giovannucci EL, Fuchs CS, Rosner BA. 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.
 66. Brenner H, Chang-Claude J, Seiler CM, Hoffmeister M. Long-term risk of colorectal cancer after negative colonoscopy. *J Clin Oncol* 2011;29:3761–7.
 67. Schoen RE. Surveillance after positive and negative colonoscopy examinations: issues, yields, and use. *Am J Gastroenterol* 2003;98:1237–46.
 68. Brenner H, Chang-Claude J, Seiler CM, Rickert A, Hoffmeister M. Protection from colorectal cancer after colonoscopy: a population-based, case-control study. *Ann Intern Med* 2011;154:22–30.
 69. Chen S, Wang W, Broman K, Katki HA, Parmigiani G. BayesMendel: an R environment for Mendelian risk prediction. Baltimore, MD: Johns Hopkins University; 2004.
 70. 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.
 71. Palomaki GE, McClain MR, Melillo S, Hampel HL, Thibodeau SN. EGAPP supplementary evidence review: DNA testing strategies aimed at reducing morbidity and mortality from Lynch syndrome. *Genet Med* 2009;11:42–65.
 72. von Holst S, Picelli S, Edler D, Lenander C, Dalén J, Hjerm F, et al. Association studies on 11 published colorectal cancer risk loci. *Br J Cancer* 2010;103:575–80.
 73. Tenesa A, Dunlop MG. New insights into the aetiology of colorectal cancer from genome-wide association studies. *Nat Rev Genet* 2009;10:353–8.
 74. Antoniou AC, Pharoah PPD, Smith P, Easton DF. The BOADICEA model of genetic susceptibility to breast and ovarian cancer. *Br J Cancer* 2004;91:1580–90.
 75. Antoniou AC, Cunningham AP, Peto J, Evans DG, Lalloo F, Narod SA, et al. The BOADICEA model of genetic susceptibility to breast and ovarian cancers: updates and extensions. *Br J Cancer* 2008;98:1457–66.
 76. Macinnis RJ, Antoniou AC, Eeles RA, Severi G, Al Olama AA, McGuffog L, 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.
 77. Vickers AJ. Prediction models: revolutionary in principle, but do they do more good than harm? *J Clin Oncol* 2011;29:2951–2.
 78. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;21:128–38.
 79. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008;27:157–72; discussion 207–12.
 80. Driver JA, Gaziano JM, Gelber RP, Lee IM, Buring JE, Kurth T. Development of a risk score for colorectal cancer in men. *Am J Med* 2007;120:257–63.
 81. Pouchet CJ, Wong N, Chong G, Sheehan MJ, Schneider G, Rosen-Shedley B, et al. A comparison of models used to predict MLH1, MSH2 and MSH6 mutation carriers. *Ann Oncol* 2009;20:681–8.
 82. Monzon JG, Cremin C, Armstrong L, Nuk J, Young S, Horsman DE, et al. Validation of predictive models for germline mutations in DNA mismatch repair genes in colorectal cancer. *Int J Cancer* 2010;126:930–9.
 83. Ma E, Sasazuki S, Iwasaki M, Sawada N, Inoue M. 10-Year risk of colorectal cancer: development and validation of a prediction model in middle-aged Japanese men. *Cancer Epidemiol* 2010;34:534–41.
 84. Imperiale TF, Wagner DR, Lin CY, Larkin GN, Rogge JD, Ransohoff DF. Using risk for advanced proximal colonic neoplasia to tailor endoscopic screening for colorectal cancer. *Ann Intern Med* 2003;139:959–65.
 85. Kim DJ, Rockhill B, Colditz GA. Validation of the Harvard Cancer Risk Index: a prediction tool for individual cancer risk. *J Clin Epidemiol* 2004;57:332–40.
 86. Emmons KM, Koch-Weser S, Atwood K, Conboy L, Rudd R, Colditz G. A qualitative evaluation of the Harvard Cancer Risk Index. *J Health Commun* 1999;4:181–93.
 87. Emmons KM, Wong M, Puleo E, Weinstein N, Fletcher R, Colditz G. Tailored computer-based cancer risk communication: correcting colorectal cancer risk perception. *J Health Commun* 2004;9:127–41.

88. Park Y, Freedman AN, Gail MH, Pee D, Hollenbeck A, Schatzkin A, et al. Validation of a colorectal cancer risk prediction model among white patients age 50 years and older. *J Clin Oncol* 2009;27:694–8.
89. Jenkins M, Southey M, Giles G, Hopper J. Rationale for, and approach to, studying modifiers of risk in persons with a genetic predisposition to colorectal cancer. *Curr Oncol Rep* 2007;9:202–7.
90. Campbell PT, Cotterchio M, Dicks E, Parfrey P, Gallinger S, McLaughlin JR. Excess body weight and colorectal cancer risk in Canada: associations in subgroups of clinically defined familial risk of cancer. *Cancer Epidemiol Biomarkers Prev* 2007;16:1735–44.
91. Campbell P, Newcomb P, Gallinger S, Cotterchio M, McLaughlin J. Exogenous hormones and colorectal cancer risk in Canada: associations stratified by clinically defined familial risk of cancer. *Cancer Causes Control* 2007;18:723–33.
92. Slattery ML, Levin TR, Ma K, Goldgar D, Holubkov R, Edwards S. Family history and colorectal cancer: predictors of risk. *Cancer Causes Control* 2003;14:879–87.
93. Win AK, Dowty JG, English DR, Campbell PT, Young JP, Winship I, et al. Body mass index in early adulthood and colorectal cancer risk for carriers and non-carriers of germline mutations in DNA mismatch repair genes. *Br J Cancer* 2011;105:162–9.
94. Watson P, Ashwathnarayan R, Lynch HT, Roy HK. Tobacco use and increased colorectal cancer risk in patients with hereditary nonpolyposis colorectal cancer (Lynch syndrome). *Arch Intern Med* 2004;164:2429–31.
95. Diergaarde B, Braam H, Vasen HF, Nagengast FM, van Muijen GN, Kok FJ, et al. Environmental factors and colorectal tumor risk in individuals with hereditary nonpolyposis colorectal cancer. *Clin Gastroenterol Hepatol* 2007;5:736–42.e1.
96. Pande M, Lynch PM, Hopper JL, Jenkins MA, Gallinger S, Haile RW, et al. Smoking and colorectal cancer in Lynch syndrome: results from the Colon Cancer Family Registry and The University of Texas M.D. Anderson Cancer Center. *Clin Cancer Res* 2010;16:1331–9.
97. Marchand LL. Combined influence of genetic and dietary factors on colorectal cancer incidence in Japanese Americans. *J Natl Cancer Inst Monogr* 1999;101–5.
98. Kune S, Kune GA, Watson LF. Case-control study of alcoholic beverages as etiological factors: the Melbourne Colorectal Cancer Study. *Nutr cancer* 1987;9:43–56.
99. Beck DE, Opelka FG, Hicks TC, Timmcke AE, Khoury DA, Gathright JB Jr. Colonoscopic follow-up of adenomas and colorectal cancer. *South Med J* 1995;88:567–70.
100. Australian Cancer Network Colorectal Cancer Guidelines Revision Committee. Guidelines for the prevention, early detection and management of colorectal cancer. Sydney, Australia: The Cancer Council Australia and Australian Cancer Network; 2005.
101. Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Ruschoff J, et al. Revised Bethesda guidelines for hereditary non-polyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst* 2004;96:261–8.
102. Vasen HF, Moslein G, Alonso A, Aretz S, Bernstein I, Bertario L, et al. Guidelines for the clinical management of familial adenomatous polyposis (FAP). *Gut* 2008;57:704–13.
103. Win AK, Cleary SP, Dowty JG, Baron JA, Young JP, Buchanan DD, et al. Cancer risks for monoallelic MUTYH mutation carriers with a family history of colorectal cancer. *Int J Cancer* 2011;129:2256–62.
104. Vogt S, Jones N, Christian D, Engel C, Nielsen M, Kaufmann A, et al. Expanded extracolonic tumor spectrum in MUTYH-associated polyposis. *Gastroenterology* 2009;137:1976–85.e10.
105. Amundadottir LT, Thorvaldsson S, Gudbjartsson DF, Sulem P, Kristjansson K, Arnason S, et al. Cancer as a complex phenotype: pattern of cancer distribution within and beyond the nuclear family. *PLoS Med* 2004;1:e65.
106. Goldgar DE, Easton DF, Cannon-Albright LA, Skolnick MH. Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. *J Natl Cancer Inst* 1994;86:1600–8.
107. Hemminki K, Chen B. Familial association of colorectal adenocarcinoma with cancers at other sites. *Eur J Cancer* 2004;40:2480–7.
108. National Cancer Institute. Colorectal cancer risk assessment tool; 2010 Jun 21 [cited 2011 Jun 28]. Available from: <http://www.cancer.gov/colorectal/cancerrisk/>.
109. Digestive Disease Institute, Medical Informatics Group, Cleveland Clinic. Cleveland Clinic colon cancer risk assessment; 2010 [cited 2011 Jun 28]. Available from: <http://digestive.ccf.org/scores/go>.