

Development and Validation of a Risk Score Predicting Risk of Colorectal Cancer

Annika Steffen¹, Robert J. MacInnis^{2,3}, Grace Joshy⁴, Graham G. Giles^{2,3}, Emily Banks^{4,5}, and David Roder¹

Abstract

Background: Quantifying the risk of colorectal cancer for individuals is likely to be useful for health service provision. Our aim was to develop and externally validate a prediction model to predict 5-year colorectal cancer risk.

Methods: We used proportional hazards regression to develop the model based on established personal and lifestyle colorectal cancer risk factors using data from 197,874 individuals from the 45 and Up Study, Australia. We subsequently validated the model using 24,233 participants from the Melbourne Collaborative Cohort Study (MCCS).

Results: A total of 1,103 and 224 cases of colorectal cancer were diagnosed in the development and validation sample, respectively. Our model, which includes age, sex, BMI, prevalent diabetes, ever having undergone colorectal cancer screening, smoking, and alcohol intake, exhibited a discriminatory accuracy of 0.73 [95% confidence interval (CI), 0.72–0.75] and 0.70 (95% CI, 0.66–0.73) using the development and validation sample, respectively. Calibration was good for both study samples. Stratified models according to colorectal cancer screening history, that additionally included family history, showed discriminatory accuracies of 0.75 (0.73–0.76) and 0.70 (0.67–0.72) for unscreened and screened individuals of the development sample, respectively. In the validation sample, discrimination was 0.68 (0.64–0.73) and 0.72 (0.67–0.76), respectively.

Conclusion: Our model exhibited adequate predictive performance that was maintained in the external population.

Impact: The model may be useful to design more powerful cancer prevention trials. In the group of unscreened individuals, the model may be useful as a preselection tool for population-based screening programs. *Cancer Epidemiol Biomarkers Prev*; 23(11); 2543–52. ©2014 AACR.

Introduction

With 1.2 million new cases in 2008, colorectal cancer ranks third of all cancers worldwide (1). It accounts for 8% of all cancer-related deaths, making it the fourth most common death from cancer (1).

More than 95% of people with colorectal cancer are estimated to benefit from curative surgery if diagnosed early (2), and screening programs have been implemented in many nations (3–5). Despite its proven effectiveness, colorectal cancer screening may have its downsides, e.g., due to investigation side effects, overdiagnosis, false-

positives, and psychologic distress (6). Also, with health-care resources increasingly in short supply, health-care services should be allocated efficiently to maximize health benefit. Greater targeting of scarce resources toward individuals at elevated risk of colorectal cancer could potentially improve cost effectiveness and health inequalities.

More than 50% of colorectal cancer cases may be linked to lifestyle factors (7), including smoking (8), lack of physical activity (9, 10), body fatness (9), alcohol (9, 10), and intake of red and processed meat or foods low in dietary fiber (9, 10). Because single risk factors are likely to cluster and interact, a model estimating each individual's risk for colorectal cancer based on multiple factors may be a valuable tool. It may inform the individuals about their risk of developing colorectal cancer, assist in determining risk-appropriate screening regimens, and may be used in medical research for designing more powerful observational studies or clinical trials (11, 12).

Only a few prediction models have been developed for incident colorectal cancer (13), based on case-control data (14), expert opinion (15), solely on men (16), an Asian population where generalizability to non-Asian settings may be in question (17) or confined to colon cancer (18). Our aim was to develop and externally validate a risk

¹University of South Australia, Division of Health Science, Adelaide, Australia. ²Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Victoria, Australia. ³Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Carlton, Victoria, Australia. ⁴National Centre of Epidemiology and Population Health, Australian National University, Canberra, Australia. ⁵The Sax Institute, Sydney, Australia.

Corresponding Author: Annika Steffen, School of Population Health, Division of Health Science, University of South Australia, Adelaide, Australia. Phone: 61-8-830-22428; Fax: 61-8-830-22794; E-mail: steay035@mymail.unisa.edu.au

doi: 10.1158/1055-9965.EPI-14-0206

©2014 American Association for Cancer Research.

score predicting absolute risk of colorectal cancer using data from two large Australian prospective studies, the 45 and Up Study and the Melbourne Collaborative Cohort Study (MCCS).

Materials and Methods

The 45 and Up Study

The Sax Institute's 45 and Up Study is a large-scale Australian cohort study designed to investigate healthy aging (19). Eligible participants were randomly selected from the Australian universal health insurance records (Medicare). A total of 267,113 men and women (53.6% women) ages ≥ 45 from the general population in New South Wales (NSW) joined the study by completing a postal questionnaire (distributed from January 2006 to December 2008) and giving written consent. Ethical approval for the study was provided by the University of NSW Human Research Ethics Committee and the Population and Health Services Research Ethics Committee.

We excluded participants with prevalent cancer (other than nonmelanoma skin cancer, $n = 55,777$), missing date of study entry ($n = 11$), with a body mass index (BMI) outside the range of 15 to 50 kg/m² ($n = 2,113$), and with invalid or most likely implausible values for physical activity and food groups under study as previously defined ($n = 11,338$; ref. 20). A total of 197,874 participants remained for analysis.

Information on predictors, including sociodemographics, medical history, body weight and height, smoking, alcohol, diet and physical activity, was derived from self-administered study questionnaires (21).

Information on cancer incidence was obtained through record linkage with the NSW Central Cancer Registry. For our analysis, the specific censoring date at which the cancer registry was considered complete was December 2008. Registry information was complemented with record data from the NSW Admitted Patient Data Collection (APDC) for the time period from January 2009 to December 2011. The APDC is a complete census of all hospital admissions and discharges in NSW and contains, among other details, the primary reason for admission. It has been shown for breast cancer that cancer diagnosis can be accurately identified using hospital data (22). The record linkage was conducted by the NSW Centre for Health Record Linkage (23).

We only considered first primary incident cases of colorectal cancer and participants were followed up from study entry to cancer diagnosis, death or follow-up termination (end December 2011), whichever came first. Incidence data were coded using the International Classification of Diseases for Oncology (ICD-O-3), with colorectal cancer comprising C18-C20 (excluding C18.1, cancers of the appendix). Proximal colon tumors included the cecum, ascending colon, hepatic flexure, and transverse colon (C18.0, 18.2–18.4). Distal colon tumors included the splenic flexure (C18.5), descending (C18.6), and sigmoid (C18.7) colon. Overlapping lesions of the colon (C18.8) and unspecified colon (C18.9) were grouped among all

colon cancers only (C18.0, C18.2–C18.9). Cancer of the rectum included tumors occurring at the rectosigmoid junction (C19) and rectum (C20). Anal canal tumors were excluded.

The MCCS

The MCCS is a prospective cohort study including 41,514 residents (41% male) of Melbourne, Australia, recruited between 1990 and 1994 and ages between 27 and 75 at baseline (24). Approval for the study was obtained from the Cancer Council Victoria's Human Research Ethics Committee and participants gave written informed consent. A structured interview was used to obtain information on sociodemographic, dietary, and lifestyle factors. Height and weight were measured according to standardized procedures. Incident cases of colorectal cancer were identified from notifications to the Victorian Cancer Registry using the same definition as in the 45 and Up Study. Risk factor data from the second follow-up (2003–2007) were used for this study to cover a similar time period for the assessment of risk factors as the development sample. Because information on alcohol was not available for the second follow-up at the time of analysis, baseline data were used. After applying the same exclusion criteria as for the development sample and additionally excluding individuals with missing information on predictors, 24,233 participants remained for analysis.

Model development

On the basis of well-established associations with colorectal cancer risk (8–10, 25), the following predictors were selected: age, BMI, sex, prevalent diabetes, previous colorectal cancer screening, first-degree relative with colorectal cancer, aspirin use, smoking status (never, former, and current), alcohol intake, cereal consumption and whole-grain bread as markers of dietary fiber intake, intake of vegetables, fruits, red meat and processed meat, and vigorous physical activity.

Missing values occurred on some predictors ranging from <1% for smoking to 13.7% for processed meat. Given that both complete-case analysis and the missing-indicator method may result in biased estimates of the associations under study and may, thus, lead to poor predictions, we used multiple imputation to handle the missing data efficiently (26, 27). Briefly, multiple imputation assumes that data are missing at random (MAR), i.e., the reason for missingness can be explained by the observed data. Missing values are sampled and replaced with a set of plausible values randomly drawn from their predicted distribution based on the other observed variables and several plausible imputed datasets are created. The results obtained from each of them are appropriately combined, fully accounting for the uncertainty caused by missing data (Rubin's rules; ref. 27). All predictor variables, the outcome variable (person-time), and the censoring variable were included in the multiple imputation procedure (27). Although some variables did not have missing

values, they might be predictive of missingness or affect the process causing missing data. Also, we further included additional variables that we considered to increase the plausibility of the MAR assumption to improve the imputation process (28), e.g., highest qualification obtained, a measure of remoteness/accessibility to services (ARIA), and an index of relative socioeconomic Advantage and Disadvantage (IRSAD), as well as frequency of chicken intake. We used the FCS (fully conditional specification) method in PROC MI in SAS, with logistic regression specified for binary and ordinal variables and regression method used for continuous variables (29, 30). Because the proportion of missing data was rather low, a total of five imputation cycles were considered reasonable to efficiently produce valid inferences. Missing data theorists indicated little or no practical benefit to using more than five to 10 imputations unless proportions of missing data are unusually high (31). A simulation study also showed that for a scenario of 10% missing data, the regression coefficients are essentially unbiased and the relative efficacy and power falloff are negligible using five compared with 100 imputations (32).

Associations of predictors with colorectal cancer were analyzed using proportional hazards regression separately by imputed dataset. On the basis of the combined estimates, weights were assigned for each predictor and the risk score was computed as a linear combination of the weighted predictors. The 5-year risk of colorectal cancer was calculated by inserting the individual risk score into the survivor function from the proportional hazards model

$$P(\text{CRC}, 5y)_i = 1 - S_M(5y)^{\exp(RS_i - RS_M)},$$

where S_M = Survivor function estimate at 5 years and at means of all predictors, RS_i = individual risk score estimated as the linear combination of weighted predictors, and RS_M = risk score estimated at means of all predictors.

Departure from the proportional hazards assumption was evaluated for all predictors based on Schoenfeld residuals. No violations were detected.

Apart from the full model defined *a priori*, we fitted a reduced model consisting of those predictors that were significantly related to colorectal cancer in the full model. We additionally computed risk scores for each colorectal cancer subtype and according to colorectal cancer screening history. We also tested for possible interactions between various predictor variables. To assess an interaction in the analysis model following multiple imputation, interaction terms need to be included in the imputation model (33). Allowing for all possible interactions, however, would make the imputation model very large (33). Following the suggestion of Wood and colleagues (33), we, therefore, performed a first imputation without interactions and derived the risk prediction model as it is presented in the results. We then explored interactions using a liberal significance level of 0.15 to allow for downward bias due to their noninclusion in the imputation model. Specifically, we tested for plausible interac-

tions of sex with BMI, smoking, physical activity, prevalent diabetes, red and processed meat intake, and also for interactions of smoking with alcohol intake and of BMI with physical activity. Because we did not detect any interactions, there was no need to subsequently update our imputation model.

Model validation

We applied the set of predictors to each of the five imputed datasets, averaged the resulting five predicted risks for each individual, and evaluated the performance of this average (34). External validation was based on the combined estimates obtained from the five imputed datasets.

Model performance was evaluated by means of discrimination and calibration. Discrimination was described by the c index for survival analysis (35, 36), which quantifies the model's ability to separate persons with longer event-free survival from those with shorter event-free survival within a given time horizon. We additionally computed the continuous net reclassification improvement [NRI(>0); refs. 37, 38] to compare the discriminatory ability of the full and reduced model. NRI(>0) values above 0.6 are considered strong, and values below 0.2, weak (39). We further calculated the integrated discrimination index (IDI), which is equivalent to the difference in discrimination slopes between the two models (37).

Calibration measures how well predicted probabilities agree with observed risks. We presented calibration at 5 years as a plot of observed proportions of events against average predicted probabilities across tenths of predicted risk.

We calculated sensitivity, specificity, positive predictive value, and negative predictive value for a range of potential cutoff points to define high-risk individuals. To find the optimal cutoff point, we used the Youden index (J), defined as $J = \text{sensitivity} + \text{specificity} - 1$ (40, 41). It allows finding the threshold for which sensitivity and specificity are maximized.

In sensitivity analyses based on the reduced model, we excluded colorectal cancer cases diagnosed within the first 2 years of follow-up. We restricted the analysis to individuals with complete information on all predictors to compare results with those obtained using multiple imputation. Finally, we evaluated sex-specific models. In all sensitivity analyses, results remained virtually unchanged.

All analyses were performed using SAS version 9.3 (SAS Inc.) and R version 2.13.1 (42, 43). For all analyses, two-sided P values were considered.

Results

In the development sample, 1,103 cases of colorectal cancer were identified during an average (\pm SD) follow-up of 3.8 ± 0.9 years [747,764 person-years (PY)]. Of these, 750 were located in the colon (456 proximal, 240 distal, and 54 unspecified) and 353 in the rectum. In the validation sample (116,455 PY), 224 individuals were diagnosed with colorectal cancer [157 colon (95 proximal, 56 distal) and 67

Table 1. General characteristics of the 45 and Up Study ($n = 197,874$) and the MCCS ($n = 24,233$)

	45 and Up Study Mean (SD) or %	MCCS Mean (SD) or %
Age at baseline (y)	61.2 (10.8)	65.7 (8.7)
% Men	42.7	38.6
Duration of follow-up (y)	3.78 (0.9)	4.83 (0.49)
BMI (kg/m ²)	27.0 (4.9)	27.3 (4.7)
First-degree relative with CRC (%)	13.2	14.4
Ever had CRC screening (%)	47.3	48.5
Prevalent diabetes (%)	8.1	7.8
Aspirin use (%)	19.6	
Smoking (%)		
Lifelong nonsmoker	58.1	62
Former smoker	34.2	32.6
Current smoker	7.7	5.4
Number of alcoholic drinks/day (median, IQR)	0.57 (0.0–1.43)	0.37 (0.0–1.33)
Vigorous activity (h/wk)		
None	59.0	
<1	16.0	NA ^a
1–3.5	16.2	
>3.5	8.8	
Diet (%)		
Red meat ≥ 5 times/wk	21.6	
Processed meat ≥ 5 times/wk	5.2	
Wholemeal bread ≥ 5 serves/wk	49.2	NA ^a
Breakfast cereals ≥ 1 serves/wk	79.5	
Vegetable ≥ 5 serves/day	34.4	
Fruits ≥ 2 serves/day	58.2	

Abbreviations: CRC, colorectal cancer; IQR, interquartile range.

^aNA because those predictors were not included in the final model.

rectum]. Distribution of predictors was broadly similar between the two samples (Table 1).

Although age, sex, BMI, prevalent diabetes, previous colorectal cancer screening, smoking, and alcohol were significantly related to colorectal cancer in our study, we did not observe a significant association with dietary factors, physical activity, family history, and aspirin use (Table 2).

The median value (range) of the risk score based on the reduced model for colorectal cancer was 4.20 (2.45–7.52) and 4.55 (2.82–7.05) in the development and validation sample, respectively. Relative risks (95% CI) of colorectal cancer by increasing score quintile for the development sample were 1.0 (reference), 1.8 (1.3–2.6), 3.2 (2.3–4.4), 6.2 (4.5–8.5), and 11.9 (8.8–16.1). Similar estimates were observed for the validation sample [1.0 (reference), 1.9 (0.9–4.0), 3.3 (1.7–6.5), 5.9 (3.1–11.2), and 9.0 (4.8–16.8), respectively].

Internal validation

The full model's c index (95% CI) was 0.73 (0.72–0.75), 0.75 (0.73–0.76), and 0.74 (0.71–0.77) for colorectal cancer,

colon, and rectum, respectively (Table 2). Discriminatory power was higher for proximal than for distal colon cancer [0.78 (0.76–0.80) vs. 0.71 (0.68–0.75), respectively, data not shown]. The restriction to significant predictors barely affected the model's discriminatory ability. This was underlined by the NRI(>0) and indicates virtually no added value of the extended model. The difference in mean predicted risk of events and nonevents for total colorectal cancer increased from 0.006007 to 0.006209 (IDI = 0.000202).

Observed and predicted risk for colorectal cancer agreed well across tenths of predicted risk with no inferiority of the reduced model (Fig. 1A and B). The reduced model was well calibrated for both distal colon and rectal cancer, whereas there was an underestimation of risk for proximal colon cancer in the middle range of risk and an overestimation of risk in the highest tenth of risk (data not shown).

External validation

Our reduced model exhibited similar discriminatory accuracy for the validation compared with development

Table 2. Full and reduced model for predicting 5-year risk of colorectal cancer in the 45 and Up Study (*n* = 197,874)

Predictor	CRC (<i>n</i> = 1103)				Colon cancer (<i>n</i> = 750)				Rectal cancer (<i>n</i> = 353)			
	Full model		Reduced model ^a		Full model		Reduced model ^a		Full model		Reduced model ^a	
	β	RR (95% CI)	β	RR (95% CI)	β	RR (95% CI)	β	RR (95% CI)	β	RR (95% CI)	β	RR (95% CI)
Age (per year)	0.063	1.07 (1.06–1.07)	0.065	1.07 (1.06–1.07)	0.072	1.08 (1.07–1.08)	0.073	1.08 (1.07–1.08)	0.043	1.04 (1.03–1.05)	0.045	1.05 (1.04–1.06)
Sex (women vs. men)	-0.283	0.75 (0.66–0.86)	-0.254	0.78 (0.68–0.88)	-0.102	0.90 (0.77–1.06)	-0.053	0.95 (0.81–1.11)	-0.691	0.50 (0.40–0.64)	-0.702	0.50 (0.39–0.62)
BMI (per kg/m ²)	0.018	1.02 (1.01–1.03)	0.020	1.02 (1.01–1.03)	0.025	1.03 (1.01–1.04)	0.027	1.03 (1.01–1.04)	0.002	1.00 (0.98–1.03)	0.006	1.01 (0.98–1.03)
Prevalent diabetes (y/no)	0.209	1.23 (1.03–1.48)	0.221	1.25 (1.04–1.50)	0.070	1.07 (0.85–1.35)	0.086	1.09 (0.87–1.37)	0.476	1.61 (1.19–2.17)	0.480	1.62 (1.20–2.18)
Family history (y/no)	0.076	1.08 (0.90–1.29)	—	—	0.081	1.08 (0.88–1.34)	—	—	0.054	1.06 (0.76–1.47)	—	—
Ever had CRC screening (y/no)	-0.536	0.58 (0.51–0.66)	-0.531	0.59 (0.52–0.67)	-0.358	0.70 (0.60–0.81)	-0.345	0.71 (0.61–0.82)	-0.943	0.39 (0.30–0.50)	-0.953	0.39 (0.30–0.49)
Aspirin use (y/no)	-0.007	0.99 (0.87–1.14)	—	—	-0.023	0.98 (0.83–1.15)	—	—	0.027	1.03 (0.80–1.32)	—	—
Smoking status												
Former smoker	0.241	1.27 (1.12–1.45)	0.240	1.27 (1.12–1.45)	0.111	1.12 (0.95–1.31)	0.109	1.11 (0.95–1.31)	0.531	1.70 (1.35–2.14)	0.536	1.71 (1.36–2.15)
Current smoker	0.254	1.29 (1.01–1.65)	0.269	1.31 (1.03–1.67)	0.179	1.20 (0.87–1.64)	0.165	1.18 (0.87–1.61)	0.401	1.49 (1.00–2.23)	0.476	1.61 (1.09–2.38)
Alcoholic drinks/day	0.077	1.08 (1.04–1.12)	0.080	1.08 (1.04–1.13)	0.099	1.10 (1.05–1.16)	0.095	1.10 (1.05–1.15)	0.035	1.04 (0.97–1.11)	0.051	1.05 (0.99–1.12)
Vigorous activity												
None	0.043	1.04 (0.87–1.25)	—	—	0.023	1.02 (0.82–1.27)	—	—	0.084	1.09 (0.80–1.48)	—	—
>1 to ≤3.5 h/wk	-0.098	0.91 (0.71–1.16)	—	—	-0.060	0.94 (0.70–1.27)	—	—	-0.173	0.84 (0.55–1.28)	—	—
>3.5 h/wk	-0.249	0.78 (0.58–1.06)	—	—	-0.358	0.70 (0.47–1.04)	—	—	-0.067	0.94 (0.58–1.51)	—	—
Red meat ≥5 times/wk	0.053	1.05 (0.91–1.21)	—	—	0.050	0.95 (0.79–1.14)	—	—	0.254	1.29 (1.02–1.64)	—	—
Processed meat ≥5 times/wk	0.146	1.16 (0.90–1.49)	—	—	0.074	1.08 (0.78–1.48)	—	—	0.248	1.28 (0.86–1.91)	—	—
Cereal intake > once/wk	-0.088	0.92 (0.78–1.08)	—	—	-0.067	0.94 (0.77–1.14)	—	—	-0.121	0.89 (0.68–1.16)	—	—
Fruits ≥2 serves/day	0.077	1.08 (0.94–1.24)	—	—	0.163	1.18 (1.00–1.27)	—	—	-0.097	0.91 (0.70–1.17)	—	—
Vegetables ≥5 serves/day	0.083	1.09 (0.95–1.24)	—	—	0.083	1.09 (0.93–1.27)	—	—	0.085	1.09 (0.85–1.39)	—	—
Wholemeal bread ≥5 serves/wk	-0.007	0.99 (0.88–1.12)	—	—	0.008	1.01 (0.87–1.17)	—	—	-0.040	0.96 (0.77–1.19)	—	—
<i>S_M</i> (5 y) ^b	0.994446	0.994411	0.996398	0.996352	0.996398	0.996352	0.996352	0.9982924	0.998277	0.9982924	0.998277	
<i>R_M</i> ^c	4.149699	4.305756	5.094325	5.177072	5.094325	5.177072	5.177072	2.102815	2.404152	2.102815	2.404152	
Internal performance												
Discrimination, c index (95% CI)	0.73 (0.72–0.75)	0.73 (0.72–0.74)	0.75 (0.73–0.76)	0.75 (0.73–0.76)	0.75 (0.73–0.76)	0.75 (0.73–0.76)	0.75 (0.73–0.76)	0.74 (0.71–0.77)	0.73 (0.71–0.76)	0.74 (0.71–0.77)	0.73 (0.71–0.76)	
NRI (>0) ^d	0.06	0.06	0.10	0.10	0.10	0.10	0.10	0.11	0.11	0.11	0.11	
IDI ^e	0.000202	0.000202	0.000218	0.000218	0.000218	0.000218	0.000218	0.000157	0.000157	0.000157	0.000157	
External performance												
Discrimination, c index (95% CI)	—	—	—	—	—	—	—	—	—	—	—	

Abbreviation: CRC, colorectal cancer.
^aAll predictors nonsignificantly associated with total colorectal cancer in full model removed.
^bSurvival function estimate at 5 years estimated at average values of all predictors.
^cRisk score at average predictor values.
^dContinuous net reclassification improvement.
^eIntegrated discrimination index.

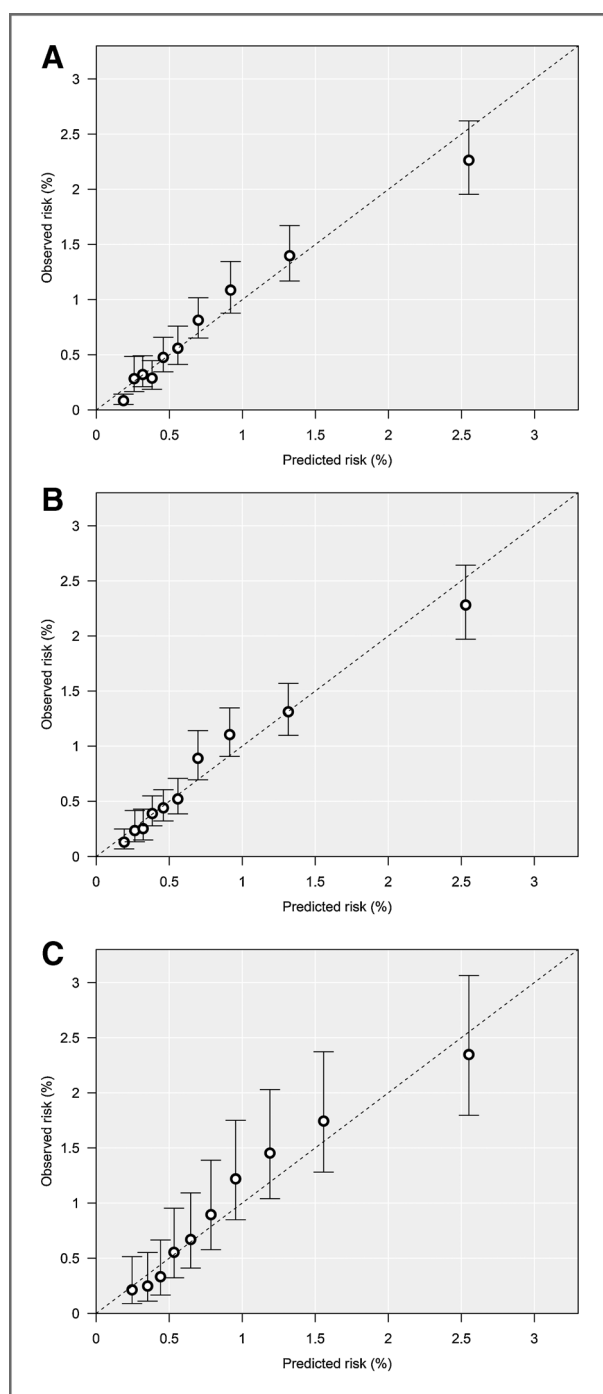


Figure 1. Calibration plot showing observed and predicted risks across tenths of predicted risk for (A) the full model in the development sample, (B) the reduced model in the development sample, and (C) the reduced model in the validation sample.

sample (Table 2). The *c* indexes (95% CI) for total colorectal cancer, colon, rectum, proximal colon and distal colon were 0.70 (0.66–0.73), 0.72 (0.68–0.76), 0.64 (0.58–0.70), 0.74 (0.70–0.78), and 0.72 (0.65–0.77), respectively. Calibration of the model for total colorectal cancer was good (Fig. 1C).

Likewise, calibration was adequate for colorectal cancer subtypes (data not shown).

Practical implications

For both populations, the Youden index suggested a risk score of ≥ 4.5 as the optimal cutoff point to define high-risk individuals based on the reduced model (Table 3). This threshold identified 80% of individuals who developed colorectal cancer during 5 years in the validation sample (sensitivity), whereas 48% of individuals not developing colorectal cancer had a risk score below this threshold (specificity). Of these individuals, the data presented here indicate that 1.0% to 1.4% will develop colorectal cancer over 5 years. It is noteworthy that for the calculation of the Youden index, sensitivity and specificity are considered equally important. This might not hold true in practice, and designation of a cutoff value should depend on the importance attached to false positives and false negatives.

Because history of colorectal cancer screening was a very strong predictor of subsequent colorectal cancer, we additionally computed stratified models according to previous colorectal cancer screening (Table 4). Interestingly, the strength of associations with many predictors appeared stronger for individuals who have never undergone colorectal cancer screening in comparison with individuals who have ever been screened. This phenomenon became particularly obvious for the risk factor family history, which was not included in the model for the total study population due to its weak and nonsignificant association with colorectal cancer. In stratified analyses, family history was associated with a significantly higher colorectal cancer risk for unscreened individuals [RR (95% CI) = 1.32 (1.04–1.69)], whereas it was not related to colorectal cancer for screened individuals [RR (95% CI) = 0.90 (0.70–1.15)]. In line with this observation, discriminatory accuracy was higher for unscreened than among screened individuals in the development sample (*c* index of 0.75 (0.73–0.76) and 0.70 (0.67–0.72), respectively). In the validation sample, discrimination was 0.68 (0.64–0.73) and 0.72 (0.67–0.77) for the unscreened and screened group, respectively.

Discussion

In this large prospective Australian study, we developed a risk score that predicts short-term risk of colorectal cancer based on personal and lifestyle factors. The model exhibited good discriminatory accuracy and calibration using an external cohort.

Strengths of our study include its prospective design, the large sample size, the inclusion of easily assessable predictors, the use of multiple imputation techniques, and the model's external validation. Interpretation of the results warrants some caution, though, as follow-up time was fairly short in respect to the long-term nature of colorectal cancer. However, results were similar after exclusion of cases occurring during the first 2 years of follow-up. Furthermore, it is not expected that middle-

Table 3. Test characteristics according to various cutoff points of the risk score based on the reduced model in the development and validation sample

Score points	Estimated risk (%)	Percentage of the population	Sensitivity (%)	Specificity (%)	Youden's index (J)	PPV (%)	NPV (%)
Development sample							
≥3.0	≥0.15	98.5	99.8	1.48	0.01	0.54	99.9
≥3.5	≥0.25	87.1	97.9	13.0	0.11	0.60	99.9
≥4.0	≥0.41	60.6	88.1	39.5	0.28	0.77	99.8
≥4.5	≥0.68	35.7	70.0	64.5	0.35	1.04	99.7
≥5.0	≥1.12	18.7	46.2	81.5	0.28	1.32	99.7
≥5.5	≥1.83	8.1	25.1	92.1	0.17	1.66	99.6
≥6.0	≥3.00	2.2	8.2	97.9	0.06	2.00	99.5
Validation sample							
≥3.0	≥0.15	99.9	100.0	0.1	0.00	0.93	100.0
≥3.5	≥0.25	95.1	99.6	4.9	0.05	0.97	99.9
≥4.0	≥0.41	78.3	95.1	21.9	0.17	1.12	99.8
≥4.5	≥0.68	52.5	79.9	47.8	0.28	1.41	99.6
≥5.0	≥1.12	27.4	52.7	72.9	0.26	1.78	99.4
≥5.5	≥1.83	9.9	22.8	90.2	0.13	2.13	99.2
≥6.0	≥3.00	2.0	7.6	98.1	0.06	3.59	99.1

NOTE: $J = [\text{sensitivity} (\%) + \text{specificity} (\%) - 100]/100$.

aged men and women dramatically change their lifestyle over time so that we can assume most predictor data to reflect longer-term lifestyle behaviors.

A prediction model aims at developing the best possible predictor rather than explaining causal associations (11). Hence, not all previously identified etiologic factors may prove to be a good predictor in such a model. In the present study, some well-described risk factors, such as aspirin use, physical activity, and family history, were not selected to remain in the final model as their added usefulness in improving the risk prediction was shown to be rather low. Consistent with previous studies, we observed a higher risk of colorectal cancer with higher BMI (9, 10), prevalent diabetes (44), current and former smoking (8), and higher intake of alcohol (9, 10). Ever having undergone colorectal cancer screening was a strong (negative) predictor for subsequent colorectal cancer risk, conferring a 40% lower risk. Randomized trials on the effectiveness of flexible sigmoidoscopy reported a reduction in colorectal cancer incidence by 18% (45). Data from the Health Professionals Follow-up study showed a risk reduction of 42% for screening endoscopy (46), which is similar to our effect size. In the present study, we did not include information on the method of colorectal cancer screening. However, in the same study population, we recently demonstrated that accounting for screening method results in largely similar risk estimates for subsequent colorectal cancer over the follow-up period of up to 5 years (47). Given this observation and the insensitivity of the c index upon inclusion of additional predictors to a relatively strong model (48), we do not expect much improvement in model performance by including the

information on screening method at this stage. As colonoscopy is assumed to be more effective in the long term, an updated prediction model covering a longer prediction time may additionally account for screening method.

Because individuals with a family history of colorectal cancer are more likely to undergo screening (71% of study participants with a family history had undergone previous colorectal cancer screening compared with 44% of participants without a family history), the inclusion of the information on previous colorectal cancer screening is likely to have attenuated the estimates for family history itself. This assumption was confirmed in analyses stratified according to screening history where family history was associated with a significantly higher colorectal cancer risk for unscreened individuals, whereas it was not related to colorectal cancer for screened individuals, suggesting that persons with a family history of colorectal cancer may counterbalance their increased risk by participating in screening. Likewise, previous screening may be a stronger negative predictor of colorectal cancer for individuals with a family history compared with individuals without a family history [RR = 0.42 (0.30–0.58) and RR = 0.62 (0.54–0.71), respectively].

The discriminatory ability of our reduced model was reasonably good and well maintained in the external validation sample that was developed entirely separately using independent methodologies. The model's performance also compares favorably with discriminatory performances of previously published colorectal cancer prediction models. In particular, Freedman and colleagues (14) presented a model predicting 10-year colorectal cancer risk based on colorectal cancer screening, polyp and

Table 4. Full and reduced model for predicting 5-year risk of colorectal cancer for screened and for unscreened individuals in the 45 and Up Study ($n = 197,874$)

Predictor	Screened individuals ($n = 93,654$ and 687 cases)			Unscreened individuals ($n = 104,220$ and 416 cases)		
	Full model		Reduced model ^a	Full model		Full model
	β	RR (95% CI)	β	RR (95% CI)	β	RR (95% CI)
Age (per year)	0.0580	1.06 (1.05–1.07)	0.0595	1.06 (1.05–1.07)	0.0654	1.07 (1.06–1.07)
Sex (women vs. men)	-0.2130	0.81 (0.65–1.00)	-0.1797	0.84 (0.68–1.03)	-0.3240	0.72 (0.61–0.85)
BMI (per kg/m ²)	0.0177	1.02 (0.99–1.04)	0.0193	1.02 (1.00–1.04)	0.0194	1.04 (1.02–1.04)
Prevalent diabetes (y/no)	0.0581	1.06 (0.79–1.49)	0.1073	1.11 (0.81–1.53)	0.2764	1.32 (1.05–1.65)
Family history (y/no)	-0.1107	0.90 (0.68–1.13)	-0.1288	0.88 (0.68–1.13)	0.2798	1.32 (1.04–1.69)
Aspirin use (y/no)	0.1055	1.10 (0.88–1.37)	—	—	-0.0769	0.93 (0.77–1.11)
Smoking status						
Former smoker	0.0236	1.04 (0.84–1.28)	0.0415	1.04 (0.84–1.29)	0.3712	1.45 (1.23–1.71)
Current smoker	0.2353	1.24 (0.79–1.96)	0.2585	1.29 (0.83–2.03)	0.2979	1.35 (1.00–1.81)
Alcoholic drinks/day	0.0697	1.06 (0.99–1.14)	0.0698	1.07 (1.00–1.15)	0.0855	1.09 (1.04–1.14)
Vigorous activity						
None	0.0572	1.06 (0.79–1.41)	—	—	0.0380	1.04 (0.83–1.30)
>1 to \leq 3.5 h/wk	-0.0224	0.94 (0.64–1.38)	—	—	-0.1152	0.89 (0.65–1.21)
>3.5 h/wk	-0.2419	0.80 (0.49–1.30)	—	—	-0.2623	0.77 (0.52–1.14)
Red meat \geq 5 times/wk	0.0924	1.09 (0.88–1.37)	—	—	0.0277	1.03 (0.85–1.24)
Processed meat \geq 5 times/wk	0.1857	1.28 (0.86–1.92)	—	—	0.0869	1.09 (0.79–1.51)
Cereal intake > once/wk	-0.2558	1.04 (0.80–1.04)	—	—	-0.0190	0.98 (0.79–1.51)
Fruits \geq 2 serves/day	0.1131	1.07 (0.85–1.34)	—	—	0.0825	1.09 (0.92–1.29)
Vegetables \geq 5 serves/day	0.0887	1.11 (0.90–1.37)	—	—	0.0667	1.07 (0.90–1.27)
Wholemeal bread \geq 5 serves/wk	-0.0069	0.99 (0.81–1.21)	—	—	-0.0051	0.99 (0.85–1.16)
S_M (5 y) ^b		0.9954152		0.9953902		0.993108
R_M ^c		3.96441		4.269461		4.544268
Internal performance						
Discrimination, c index (95% CI)		0.70 (0.67–0.73)		0.70 (0.67–0.72)		0.75 (0.73–0.76)
NRI (>0) ^d			0.13			0.09
IDI ^e			0.00027			0.00020
External performance						
Discrimination, c index (95% CI)				0.72 (0.67–0.77)		0.68 (0.64–0.73)

^aAll predictors nonsignificantly associated with total colorectal cancer in full model removed. Please note that, in contrast with the reduced model in the total study population, family history is included here.

^bSurvival function estimate at 5 years estimated at average values of all predictors.

^cRisk score at average predictor values.

^dContinuous net reclassification improvement.

^eIntegrated discrimination index.

family history, BMI, aspirin use, smoking and consumption of vegetables; external validation in the NIH-AARP study exhibited a discriminatory accuracy of 0.61 (49). The Physician's Health Study model that includes age, BMI, smoking, and alcohol yielded a c statistic of 0.70 over a 20-year time period (16). The discriminatory ability of a model predicting 10-year colorectal cancer risk using a Japanese cohort was 0.70 (0.68–0.72) for the development cohort and 0.64 (0.61–0.67) for a Japanese external validation cohort (17).

The stronger associations with several predictors and the higher predictive accuracy of the model found for individuals who have never undergone colorectal cancer screening compared with individuals with a history of colorectal cancer screening suggests that a previous colorectal cancer screening may counterbalance the effect and consequently the predictive ability of the other included risk factors.

In terms of practical application, our prediction model may help to design more powerful clinical trials by enriching the number of observed events. The additional model for the subgroup of unscreened individuals may be valuable in defining inclusion criteria for risk-based colorectal cancer screening programs (12). Invitation to colorectal cancer screening is commonly based on age criteria (3–5). Although age is a major risk factor of cancer, cancer risk is also affected by other determinants, and it has been suggested to replace the age criterion by a more general risk criterion (12). The use of prediction models as a preselection tool in population-based screening programs may improve the benefit-harms ratio of screening and assist in focusing scarce resources.

Before any implementation, the effect of using the model in the envisaged field of application, including a careful evaluation of health outcomes and cost effectiveness, needs to be quantified (50). In cancer risk prediction, only the Gail model for breast cancer has currently reached the phase of impact analysis (12). For the Gail model, discriminatory accuracies between 0.58 and 0.67 have been reported (51–54).

In conclusion, we have developed a risk score that predicts short-term risk of colorectal cancer and have demonstrated that it performs well in an independent

population. The model may be useful in trial-based research. As a re-estimated version for individuals who have never undergone screening, it may be used as a preselection tool for population-based screening programs.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: A. Steffen, G.G. Giles

Development of methodology: A. Steffen, R.J. MacInnis

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): G.G. Giles, E. Banks, D. Roder

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): A. Steffen, R.J. MacInnis, G. Joshy, D. Roder

Writing, review, and/or revision of the manuscript: A. Steffen, R.J. MacInnis, G. Joshy, G.G. Giles, E. Banks, D. Roder

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): G.G. Giles

Study supervision: D. Roder

Acknowledgments

The Melbourne Collaborative Cohort Study was made possible by the contribution of many people, including the original investigators and the diligent team who recruited the participants, and who continue working on follow-up. The authors thank the many thousands of Melbourne residents who continue to participate in the 45 and Up Study. The authors also acknowledge the support of the Centre for Health Record Linkage.

Grant Support

This research was completed using data collected through the 45 and Up Study (www.saxinstitute.org.au). The 45 and Up Study is managed by the Sax Institute in collaboration with major partner Cancer Council NSW; and partners: the National Heart Foundation of Australia (NSW Division); NSW Ministry of Health; *beyondblue*; Ageing, Disability, and Home Care, Department of Family and Community Services; the Australian Red Cross Blood Service; and UnitingCare Ageing.

This work was supported by infrastructure from the Cancer Council Victoria and grants from the National Health and Medical Research Council of Australia 209057 and 251533.

A. Steffen was supported by a scholarship from the German Research Foundation (DFG). E. Banks is supported by the National Health and Medical Research Council of 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 February 24, 2014; revised July 11, 2014; accepted July 29, 2014; published OnlineFirst August 3, 2014.

References

1. Ferlay J, Shin H, Bray F, Forman D, Mathers C, Parkin DM. GLOBOCAN 2008 v1.2, Cancer Incidence and Mortality Worldwide: IARC Cancer-Base No. 10. 2010 [cited 2011 09/12]; Available from: <http://globocan.iarc.fr>
2. Bustin SA, Murphy J. RNA biomarkers in colorectal cancer. *Methods* 2013;59:116–25.
3. Australian Institute of Health and Welfare. National Bowel Cancer Screening Program monitoring report: phase 2, July 2008–June 2011. *Cancer Series no. 66. Cat. No. CAN 62. Canberra: AIHW; 2012.*
4. Levin B, Lieberman DA, McFarland B, Andrews KS, Brooks D, Bond J, 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 Colo-rectal Cancer, and the American College of Radiology. *Gastroenterology* 2008;134:1570–95.
5. von Karsa L, Anttila A, Ronco G, et al. Cancer Screening in the European Union. Report on the implementation of the Council Recommendation on cancer screening. Lyon: International Agency for Research on Cancer. [cited 2012 Oct. 30]. Available from: http://ec.europa.eu/health/ph_determinants/genetics/documents/cancer_screening.pdf; 2008.
6. Bretthauer M, Kalager M. Principles, effectiveness and caveats in screening for cancer. *Br J Surg* 2013;100:55–65.
7. Parkin DM, Boyd L, Walker LC. 16. The fraction of cancer attributable to lifestyle and environmental factors in the UK in 2010. *Br J Cancer* 2011;105 Suppl 2:S77–81.

8. Botteri E, Iodice S, Bagnardi V, Raimondi S, Lowenfels AB, Maisonneuve P. Smoking and colorectal cancer: a meta-analysis. *JAMA* 2008;300:2765–78.
9. World Cancer Research Fund, American Institute for Cancer Research. Food, Nutrition, Physical Activity, and the Prevention of Cancer: A Global Perspective. Washington DC: AICR; 2007.
10. World Cancer Research Fund, American Institute for Cancer Research. Continuous Update Project. Available from: http://www.dietandcancer-report.org/cup/current_progress/colorectal_cancer.php (last access: 04/10/2012); 2011.
11. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ* 2009;338:b606.
12. Stegeman I, Bossuyt PM. Cancer risk models and preselection for screening. *Cancer Epidemiol* 2012;36:461–9.
13. Win AK, Macinnis RJ, Hopper JL, Jenkins MA. Risk prediction models for colorectal cancer: a review. *Cancer Epidemiol Biomarkers Prev* 2012;21:398–410.
14. 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.
15. Colditz GA, Atwood KA, Emmons K, Monson RR, Willett WC, Trichopoulos D, et al. Harvard report on cancer prevention volume 4: Harvard Cancer Risk Index. Risk Index Working Group, Harvard Center for Cancer Prevention. *Cancer Causes Control* 2000;11:477–88.
16. 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.
17. 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* 2011;34:534–41.
18. 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.
19. 45 and Up Study Collaborators Banks E, Redman S, Jorm L, Armstrong B, Bauman A, et al. Cohort profile: the 45 and up study. *Int J Epidemiol* 2008;37:941–7.
20. Sax Institute. 45 and Up Study Technical Note 1: Missing or Invalid Values; 2013. [cited 2014 June 22]. Available from: <https://www.saxinstitute.org.au/wp-content/uploads/Technical-Note-missing-or-invalid-values.pdf>.
21. Sax Institute. The Baseline Questionnaires. [cited 2014 June 22]. Available from: <https://www.saxinstitute.org.au/our-work/45-up-study/questionnaires/>.
22. Kemp A, Preen DB, Saunders C, Holman CD, Bulsara M, Rogers K, et al. Ascertaining invasive breast cancer cases; the validity of administrative and self-reported data sources in Australia. *BMC Med Res Methodol* 2013;13:17.
23. Centre for Health Record Linkage. [cited 2012 Nov. 17]. Available from: <http://www.cherel.org.au/>.
24. Giles GG, English DR. The Melbourne Collaborative Cohort Study. *IARC Sci Publ* 2002;156:69–70.
25. Bosetti C, Rosato V, Gallus S, Cuzick J, La Vecchia C. Aspirin and cancer risk: a quantitative review to 2011. *Ann Oncol* 2012;23:1403–15.
26. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol* 2006;59:1087–91.
27. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009;338:b2393.
28. van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat Med* 1999;18:681–94.
29. Allison PD. Imputation of categorical variables with PROC MI. *Proceedings* 2005, 113–30, pp. 1–14.
30. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat Methods Med Res* 2007;16:219–42.
31. Schafer JL. Multiple imputation: a primer. *Stat Methods Med Res* 1999;8:3–15.
32. Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prev Sci* 2007;8:206–13.
33. Wood AM, White IR, Royston P. How should variable selection be performed with multiply imputed data? *Stat Med* 2008;27:3227–46.
34. Vergouwe Y, Royston P, Moons KG, Altman DG. Development and validation of a prediction model with missing predictor data: a practical approach. *J Clin Epidemiol* 2010;63:205–14.
35. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361–87.
36. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat Med* 2004;23:2109–23.
37. 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.
38. Pencina MJ, D'Agostino RB Sr, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med* 2011;30:11–21.
39. Pencina MJ, D'Agostino RB, Pencina KM, Janssens AC, Greenland P. Interpreting incremental value of markers added to risk prediction models. *Am J Epidemiol* 2012;176:473–81.
40. Bewick V, Cheek L, Ball J. *Statistics review 13: receiver operating characteristic curves*. *Crit Care* 2004;8:508–12.
41. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;3:32–5.
42. Johnston D, Gong G. *rmap: Risk Model Assessment Package*. R package version 0.01-02; 2011.
43. R Development Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, ISBN 3-900051-07-0; 2011. Available from: <http://www.R-project.org/>
44. Deng L, Gui Z, Zhao L, Wang J, Shen L. Diabetes mellitus and the incidence of colorectal cancer: an updated systematic review and meta-analysis. *Dig Dis Sci* 2012;57:1576–85.
45. Elmunzer BJ, Hayward RA, Schoenfeld PS, Saini SD, Deshpande A, Waljee AK. Effect of flexible sigmoidoscopy-based screening on incidence and mortality of colorectal cancer: a systematic review and meta-analysis of randomized controlled trials. *PLoS Med* 2012;9:e1001352.
46. Kavanagh AM, Giovannucci EL, Fuchs CS, Colditz GA. Screening endoscopy and risk of colorectal cancer in United States men. *Cancer Causes Control* 1998;9:455–62.
47. Steffen A, Weber M, Roder D, Banks E. Colorectal cancer screening and subsequent incidence of colorectal cancer: results from the 45 and Up Study. *Medical Journal of Australia* (accepted for publication).
48. Cook NR. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. *Clin Chem* 2008;54:17–23.
49. 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.
50. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* 2012;98:691–8.
51. Tice JA, Cummings SR, Ziv E, Kerlikowske K. Mammographic breast density and the Gail model for breast cancer risk prediction in a screening population. *Breast Cancer Res Treat* 2005;94:115–22.
52. MacKarem G, Roche CA, Hughes KS. The effectiveness of the Gail model in estimating risk for development of breast cancer in women under 40 years of age. *Breast J* 2001;7:34–9.
53. Chen J, Pee D, Ayyagari R, Graubard B, Schairer C, Byrne C, et al. Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. *J Natl Cancer Inst* 2006;98:1215–26.
54. Gail MH. Discriminatory accuracy from single-nucleotide polymorphisms in models to predict breast cancer risk. *J Natl Cancer Inst* 2008;100:1037–41.