



Practice of Epidemiology

A Comprehensive Model of Colorectal Cancer by Risk Factor Status and Subsite Using Data From the Nurses' Health Study

Esther K. Wei*, Graham A. Colditz, Edward L. Giovannucci, Kana Wu, Robert J. Glynn, Charles S. Fuchs, Meir Stampfer, Walter Willett, Shuji Ogino, and Bernard Rosner

* Correspondence to Dr. Esther K. Wei, California Pacific Medical Center Research Institute, 475 Brannan Street, San Francisco, CA 94107 (e-mail: esther.wei@channing.harvard.edu).

Initially submitted August 6, 2015; accepted for publication March 31, 2016.

We expanded and updated our colon cancer risk model to evaluate colorectal cancer (CRC) and whether subsite-specific risk models are warranted. Using data from 1980–2010 for 90,286 women enrolled in the Nurses' Health Study, we performed competing-risks regression and tests for subsite heterogeneity (proximal colon: $n = 821$; distal colon: $n = 521$; rectum: $n = 376$). Risk factors for CRC were consistent with those in our colon cancer model. Processed meat consumption was associated with a higher risk of distal (hazard ratio (HR) = 1.45; $P = 0.02$) but not proximal (HR = 0.95; $P = 0.72$) colon cancer. Smoking was associated with both colon (HR = 1.21) and rectal (HR = 1.27) cancer and was more strongly associated with proximal (HR = 1.31) than with distal (HR = 1.04) colon cancer ($P = 0.029$). We observed a significant trend of cancer risk for smoking in subsites from the cecum (HR = 1.41) to the proximal colon (excluding the cecum; HR = 1.27) to the distal colon (HR = 1.04; P for trend = 0.040). The C statistics for colorectal ($C = 0.607$), colon ($C = 0.603$), and rectal ($C = 0.639$) cancer were similar, although C was slightly higher for rectal cancer. Despite evidence for site-specific differences for several risk factors, overall our findings support the application of risk prediction models for colon cancer to CRC.

colorectal cancer; incidence; rectal cancer; risk; risk factors; risk prediction model; women

Abbreviations: BMI, body mass index; CRC, colorectal cancer; FFQ, food frequency questionnaire; HR, hazard ratio; MET, metabolic equivalent; PMH, postmenopausal hormone.

Cancer incidence models have flourished in the context of breast cancer (1–6) due to the established role of reproductive factors and known variation in hormone levels throughout a woman's life (7–9). Despite established risk factors for colorectal cancer (CRC), few models comprehensively evaluate CRC risk factors simultaneously and evaluate the interplay of these risk factors in the cumulative incidence of CRC over time (10–12). Moreover, the literature lacks a comprehensive model of CRC incidence that evaluates the associations for risk factors by specific subsites within the colorectum. Increasing evidence supports variation in the pathological characteristics of CRCs along the colorectal segments, possibly due not only to differential embryonic origins but also to exposure to varying bowel contents (including microbiota) and host immune

response (13–19). We have expanded and updated our Rosner-Wei model of colon cancer incidence (20) to include both colon and rectal cancer, additional follow-up time, and tests for differential associations by subsite.

METHODS

Study population

We used data from the Nurses' Health Study, a prospective study of a cohort of 121,701 female nurses in the United States, who were enrolled in 1976 and have completed biennial questionnaires updating information regarding lifestyle factors, medication use, disease endpoints, and family history. Dietary data have been collected via self-administered

and validated food frequency questionnaire (FFQ) every 2–4 years since 1980. Our follow-up began at the return of the 1980 questionnaire and ended on the date of return of the last questionnaire within our study period (June 1, 2010), the development of any cancer, or death, whichever occurred first. We excluded women who were missing information on date of birth, height, or weight at age 18 years and those who reported a cancer diagnosis at or before baseline (1980), leaving 1,759 colorectal, 1,345 colon, and 380 rectal cancer cases. (The total numbers of colon and rectal cancers do not sum to 1,759 due to incomplete data.) We excluded an additional 106 participants (including 3 colon cancer and 4 rectal cancer cases) based on their outlier values for folate intake, alcohol intake, meat intake, or physical activity using the generalized “extreme Studentized deviate” many-outlier detection method (21). Proximal cancers included those in the cecum, ascending colon, or transverse colon; distal cancers occurred in the descending or sigmoid colon. Our final sample included 1,342 colon cancers (proximal: $n = 821$, distal: $n = 521$) and 376 rectal cancers.

Assessment of risk factors

Dietary data were collected via validated FFQs in 1980, 1984, and 1986 and thereafter every 4 years (22). We included total dietary intake of folate, calcium, alcohol, and red and processed meats, and we updated intake data every 2–4 years. If a participant was missing a response for a specific questionnaire year for an item that was not missing for the previous cycle, we “carried forward” the prior nonmissing value for 1 cycle. Any remaining missing values were assigned the median value for all subjects, averaged over the entire follow-up period. We estimated beta coefficients for nonprocessed red meat and processed meat using individual foods reported on the FFQ. Red meat included hamburger, beef/pork/lamb as a mixed dish, and beef/pork/lamb as a main dish, and processed meat included hot dogs, processed meat, processed meat sandwiches, and bacon.

Nondietary variables included family history of colon or rectal cancer in a first-degree relative, cumulative smoking history, body mass index (BMI), leisure-time physical activity, height, aspirin use, history of colorectal screening by endoscopy, and postmenopausal hormone (PMH) treatment. Colon or rectal cancer in a first-degree relative was assessed in 1982, 1988, 1992, and thereafter every 4 years. For smoking, we derived a total cumulative pack-years variable. We did not add pack-years to the cumulative total for missing smoking information. Height was reported once in 1976 and held constant throughout follow-up. Current weight was reported on every questionnaire starting in 1976. The validity of self-reported current weight is high (23). Height and weight were used to calculate BMI ($\text{weight (kg)/height (m)}^2$), and if weight was missing, we carried forward the most recent value if the previous questionnaire value was not missing. Using detailed physical activity information on type and duration of specific activities, collected every 4 years starting in 1986, we calculated total metabolic equivalent (MET) hours/week. Before 1986, we were unable to calculate MET-hours/week and assigned zero for all participants for that time period.

Missing values were assigned the median value. A simple yes/no question on aspirin use (available starting in 1980) and a question on frequency of use (available starting in 1984) were used to derive a cumulative average number of aspirin tablets per week (continuous variable). Participants with missing values were assumed to be nonusers. Participants were asked whether they had undergone a “colonoscopy or sigmoidoscopy” (endoscopy) in 1988, 1990, 1992, and every 2 years after that. On the 1990 questionnaire, participants reported their history of endoscopy for the period of 1980–1990. Missing values were assumed to mean that participants did not undergo screening during that cycle. When a woman reported screening endoscopy, we assigned her 2 questionnaire cycles (4 years) of screening “coverage,” starting from the age at which she reported being screened. Although 10 years is the recommended colonoscopy interval, in the early follow-up period we did not have information on type of endoscopy, so our intent was to approximate the 5-year recommended interval for sigmoidoscopy. Participants reported menopausal status and PMH use on every questionnaire; women missing data for hormone use were assumed to be nonusers. Web Table 1 (available at <http://aje.oxfordjournals.org/>) shows the overall distribution of the risk factors in our model at baseline and for 4 additional questionnaire cycles.

For the dietary variables and BMI, we created a “calendar” variable that reflected individual intakes and BMI for every age during follow-up, starting at age in 1980. We did this by using each participant’s age and reported intakes and BMI on each questionnaire cycle. For ages between questionnaire cycles, participants were assigned the most recently reported value. For example, for a woman who was 40 years old in 1980, her reported intake of calcium in 1980 was used as her calcium intake at ages 40, 41, 42, and 43 years (minus a mean centering value). When she completed the 1984 FFQ, she would be 44 years old; thus her reported intake on that FFQ would be assigned as her intake at ages 44 and 45 years. In 1986, when she completed another FFQ, her reported intake would be assigned as her intake at ages 46 and 47 years, and so forth for the entire follow-up period. We summed the intakes from age 30 years to the end of follow-up for each participant (updating the sum at each questionnaire cycle), and we entered this value into the model as a continuous variable. For a woman with a folate intake above the median throughout follow-up, the sum value would be positive, whereas a woman with folate intake below the median would have a negative-sum folate value. Thus, beta coefficients for these variables are interpreted as the association of that variable for a given contrast per year multiplied by the number of years since 1980 (e.g., folate intake, 600 μg vs. 200 μg per year \times 30 years).

Statistical analysis

We fitted an extension of the Cox proportional hazards model applied to the setting of competing-risks survival (24) using PROC PHREG with age as the time metameter (SAS, version 9.3; SAS Institute, Inc., Cary, North Carolina). Briefly, competing-risk modeling involves data augmentation where each subject has a separate observation for each outcome, with each observation having the same follow-up time

but with a separate indicator variable for each subsite. Similar to previous studies, we stratified on event subtype to estimate separate associations of each risk factor with each outcome under a proportional hazard assumption in a single model (25–27). The hazard ratios were calculated as $\exp(\beta)$ for dichotomous variables and hazard ratio = $\exp(\beta) \times$ contrast for continuous variables. For variables that were calculated per year, we used $\exp(\beta) \times$ (contrast in risk factor) \times (years at that exposure level). We report Wald tests for differential associations at the various subsites for each risk factor and P values for pairwise comparisons (all P values were 2-sided). We previously tested for interactions, and none of the interaction terms were statistically significant (20). Because the goal of this analysis was to evaluate our previous model for subsite differences, and in order to create a parsimonious model, we did not include interactions in this analysis. Using the TEST statement in PROC PHREG, we tested for linear trends across the colon (trend 1) and colorectum (trend 2). We multiplied the beta coefficient for each subsite by a weighted average of relative (ordinal) distance from the anus to account for the relative distances from the anus (sample calculations in the table footnotes). The C statistics were calculated within 10-year age groups, and an overall C statistic was obtained by calculating a weighted average of the age-specific C statistics, where the weights were equal to 1 divided by the estimated variance (28). We compared C statistics from alternative prediction models while controlling for age and tested for heterogeneity of the C statistics by subsite (29). We calculated survival probabilities for ages 50–70 years for various combinations of risk factors, using the BASELINE option of SAS PROC PHREG; results are shown in figures below and Web Figures 1 and 2.

Description of the model

For this analysis, our final model was

$$\begin{aligned} \log h(t) = & \log h_0(t) + \beta_1 \text{AGE} + \beta_2 \text{FHX} \\ & + \beta_{3A} \sum_{j=180}^{t-1} (\text{RMEAT}_j - 1.0) + \beta_{3B} \sum_{j=180}^{t-1} (\text{PMEAT}_j - 1.0) \\ & + \beta_4 \sum_{j=180}^{t-1} (\text{FOL}_j - 390) + \beta_5 \sum_{j=180}^{t-1} (\text{CALC}_j - 800) \\ & + \beta_6 \sum_{j=180}^{t-1} (\text{SMK}_j - 12) + \beta_7 \sum_{j=180}^{t-1} (\text{BMI}_j - 26) \\ & + \beta_8 \sum_{j=186}^{t-1} (\text{ACT}_j - 11) + \beta_9 \sum_{j=180}^{t-1} (\text{HGT}_j - 64.5) \\ & + \beta_{10} \sum_{j=180}^{t-1} \text{ALC}_j + \beta_{11} \sum_{j=180}^{t-1} \text{ASP}_j + \beta_{12} \sum_{j=180}^{t-1} \text{SCREEN}_j \\ & + \beta_{13} \text{PMH}_{\text{cur},j} + \beta_{14} \text{PMH}_{\text{past},j} \end{aligned}$$

where

- $t80$ = age in 1980 and t = age at time t and
- FHX = 1 if family history of colon or rectal cancer, = 0 otherwise;
- RMEAT $_j$ = intake of red meat at age j (servings/day);
- PMEAT $_j$ = intake of processed meat at age j (servings/day);

- FOL $_j$ = total intake of folate at age j ($\mu\text{g}/\text{day}$);
- CALC $_j$ = total intake of calcium at age j (mg/day);
- SMK $_j$ = total cumulative pack-years smoked;
- BMI $_j$ = body mass index at age j ;
- ACT $_j$ = physical activity in MET-hours/week at age j ;
- HGT = adult attained height (inches);
- ALC $_j$ = intake of alcohol at age j (g/day);
- ASP $_j$ = number of aspirin tablets per week at age j ;
- SCREEN $_j$ = cumulative number of years of endoscopy screening coverage by age j ;
- PMH $_{\text{cur},j}$ = 1 if current PMH use at age j , = 0 otherwise;
- PMH $_{\text{past},j}$ = 1 if past PMH use at age j , = 0 otherwise.

The interpretation of each of the coefficients is the hazard ratio for β_1 : a 1-year increase in age; β_2 : those who reported a family history of colon or rectal cancer; β_{3A} : 1 serving per day of red meat, per year; β_{3B} : 1 serving per day of processed meat, per year; β_4 : 1 μg total folate intake per day, per year; β_5 : 1 mg total calcium intake per day, per year; β_6 : 1 pack-year of cigarette smoking; β_7 : 1 BMI-year (i.e., a 1-unit change in BMI per year); β_8 : 1 MET-hour of physical activity per week, per year; β_9 : 1 inch in height, per year; β_{10} : 1 additional gram per day of alcohol intake, per year; β_{11} : 1 tablet of aspirin used per week, per year; β_{12} : 1 year of screening coverage by endoscopy (sigmoidoscopy or colonoscopy); and β_{13} and β_{14} : current and past PMH use (vs. never use).

RESULTS

Table 1 shows the age-standardized distribution of selected variables as reported in 1980. Women with CRC were older, were slightly taller, had a higher average BMI, had a higher total number of pack-years smoked, and included a higher percentage of women who had never used PMH and who had a positive family history of colon or rectal cancer. Women with CRC also had lower folate and calcium intakes, were less physically active, took less aspirin, and were less likely to report screening endoscopy.

Table 2 shows the hazard ratios, 95% confidence intervals, and P values for colorectal, colon, and rectal cancer. Only the 2 meat variables and past PMH use were not statistically significantly associated with CRC, and alcohol intake had a borderline association ($P = 0.082$). Age, smoking history, and BMI were associated with a statistically significantly higher risk of colorectal and, separately, colon and rectal cancer risk (Web Figures 1A and 1B). Family history was statistically significantly associated with colorectal and colon cancer (hazard ratio (HR) for colon cancer = 1.50; $P \leq 0.0001$) but not rectal cancer (HR = 1.26; $P = 0.095$), although this difference was not statistically significant (P for colon vs. rectal = 0.27). The point estimates for colorectal, colon, and rectal cancer were identical for height (HR = 1.24; P for colon vs. rectal = 0.97).

Among dietary factors, folate intake was statistically significantly inversely associated with CRC (HR = 0.83; $P = 0.004$) (Web Figure 2A) and colon cancer (HR = 0.85; $P = 0.021$), but its association with rectal cancer was borderline (HR = 0.77, 95% confidence interval: 0.58, 1.02;

Table 1. Age-Standardized Characteristics of Colorectal Cancer Cases and Noncases at Baseline in the Nurses' Health Study, 1980

Characteristic	Case Group (n = 1,759)		Noncase Group (n = 88,527)	
	Mean (SD)	%	Mean (SD)	%
Age, years ^a	49.6 (6.5)		45.9 (7.2)	
Red meat intake, servings/day ^b	0.80 (0.47)		0.78 (0.48)	
Processed meat intake, servings/day ^c	0.38 (0.39)		0.37 (0.38)	
Height, inches	64.7 (2.4)		64.5 (2.4)	
Folate intake, µg/day	341 (214)		364 (263)	
Calcium intake, mg/day	714 (297)		732 (310)	
Physical activity, MET-hours/week ^d				
BMI ^e	24.8 (4.7)		24.3 (4.4)	
Alcohol intake (all participants), g/day	6 (11)		6 (10)	
Alcohol intake (among alcohol drinkers), g/day	10 (13)		9 (11)	
Smoking history (all participants), pack-years	12 (16)		11 (15)	
Smoking history (among smokers), pack-years	22 (17)		20 (16)	
Aspirin use (all participants), tablets/week	2.1 (4.3)		2.7 (4.9)	
Aspirin use (among users), tablets/week	6.2 (5.4)		6.8 (5.6)	
PMH use (among postmenopausal women)				
Never use		64		57
Current use		13		19
Past use		23		24
Postmenopausal		43		43
Family history of colon or rectal cancer ^f		12		8
Report of sigmoidoscopy or colonoscopy		8		10

Abbreviations: BMI, body mass index; MET, metabolic equivalent; PMH, postmenopausal hormone; SD, standard deviation.

^a Not age-standardized.

^b Red meat included the following specific food items: hamburger, beef/pork/lamb as a mixed dish, and beef/pork/lamb as a main dish.

^c Processed meat variable included the following specific food items: hot dogs, processed meat, processed meat sandwiches, and bacon.

^d Physical activity level in MET-hours/week was calculated starting in 1986.

^e BMI was calculated as weight (kg)/height (m)².

^f Family history of colon or rectal cancer in a first-degree relative.

$P = 0.071$); the test for heterogeneity was not significant (P for colon vs. rectal = 0.55). The inverse association with calcium intake was similar in magnitude by subsite and was statistically significant for colon cancer (HR = 0.80; $P = 0.0004$) but not for rectal cancer (HR = 0.89; $P = 0.36$ and P for colon vs. rectal = 0.43) (Web Figure 2B).

Intakes of red meat (HR for colon = 1.02 and HR for rectal = 0.99) and processed meat (HR for colon = 1.12 and HR for rectal = 1.13) were not associated with the risk of colon or rectal cancer (P for colon vs. rectal = 0.88 for red meat and 0.98 for processed meat). Alcohol intake was not significantly associated with either colon cancer (HR = 1.08) or rectal (HR = 1.33; P for colon vs. rectal = 0.28) cancer.

Current PMH use was inversely associated with risk of CRC (HR = 0.87; $P = 0.026$), and the association did not vary by subsite (HR for colon = 0.88 and HR for rectal = 0.88; P for colon vs. rectal = 0.97). Past PMH use

was not statistically significantly associated with any site (HR for colorectal = 0.93, HR for colon = 0.99, and HR for rectal = 0.79; P for colon vs. rectal = 0.12).

Aspirin intake was associated with a 22% lower risk of CRC for those who reported taking 7 aspirin tablets per week compared with those who reported no aspirin use, and this did not vary by subsite (HR for colon = 0.76 and HR for rectal = 0.80; P for colon vs. rectal = 0.76).

Reporting 21 MET-hours/week of physical activity was associated with a statistically significant reduction in risk of CRC (HR = 0.61; $P < 0.0001$) (Web Figure 1C) and colon cancer (HR = 0.55; $P < 0.0001$), but the association was attenuated for rectal cancer (HR = 0.89; $P = 0.63$ and P for colon vs. rectal = 0.067).

Screening coverage was associated with a statistically significantly lower risk of CRC (HR = 0.74; $P < 0.0001$). However, endoscopic screening was the only variable that

Table 2. Hazard Ratios for Incident Colon Cancer and Rectal Cancer (Competing-Risks Model) in the Nurses' Health Study, 1980–2010

Risk Factor	Colorectal Cancer (n = 1,759 ^a)			Colon Cancer (n = 1,342)			Rectal Cancer (n = 376)			P Value Colon vs. Rectal
	HR ^b	95% CI	P Value	HR ^b	95% CI	P Value	HR ^b	95% CI	P Value	
Age (years; 60 vs. 50)	1.81	1.70, 1.92	<0.0001	1.85	1.72, 1.98	<0.0001	1.71	1.50, 1.96	<0.0001	0.33
Family history of colon or rectal cancer (yes vs. no) ^c	1.45	1.29, 1.63	<0.0001	1.50	1.31, 1.71	<0.0001	1.26	0.96, 1.66	0.095	0.27
Red meat intake (servings/day per year; 1 vs. 0)	1.01	0.87, 1.18	0.87	1.02	0.86, 1.21	0.81	0.99	0.70, 1.40	0.96	0.88
Processed meat intake (servings/day per year; 1 vs. 0)	1.11	0.92, 1.33	0.29	1.12	0.91, 1.39	0.28	1.13	0.75, 1.71	0.57	0.98
Folate intake (μg/day per year; 600 vs. 200)	0.83	0.74, 0.95	0.004	0.85	0.74, 0.97	0.021	0.77	0.58, 1.02	0.071	0.55
Smoking history (total pack-years; 40 vs. 0)	1.20	1.10, 1.31	<0.0001	1.21	1.09, 1.33	0.0002	1.27	1.05, 1.53	0.013	0.64
BMI (units per year; 30 vs. 20) ^d	1.37	1.19, 1.57	<0.0001	1.32	1.13, 1.56	0.0006	1.50	1.10, 2.04	0.011	0.49
Physical activity level (MET-hours/week per year; 21 vs. 2)	0.61	0.48, 0.76	<0.0001	0.55	0.42, 0.71	<0.0001	0.89	0.56, 1.41	0.63	0.067
Height (inches per year; 67 vs. 61)	1.24	1.09, 1.41	0.001	1.24	1.07, 1.44	0.003	1.24	0.93, 1.65	0.15	0.97
Alcohol (g/day per year; 30 vs. 0)	1.15	0.98, 1.34	0.082	1.08	0.90, 1.29	0.39	1.33	0.96, 1.85	0.088	0.28
Aspirin use (tablets per week per year; 7 vs. 0)	0.78	0.70, 0.86	<0.0001	0.76	0.67, 0.86	<0.0001	0.80	0.63, 1.01	0.060	0.76
Endoscopic screening (yes vs. no; 20 years vs. 0) ^e	0.74	0.67, 0.83	<0.0001	0.82	0.73, 0.92	0.0005	0.48	0.36, 0.64	<0.0001	0.0008
PMH use (vs. never users)										
Current use	0.87	0.76, 0.98	0.026	0.88	0.76, 1.01	0.075	0.88	0.68, 1.15	0.35	0.97
Past use	0.93	0.83, 1.05	0.24	0.99	0.87, 1.13	0.89	0.79	0.61, 1.02	0.074	0.12
Calcium intake (mg/day per year; 1,000 vs. 500)	0.82	0.73, 0.91	0.0002	0.80	0.71, 0.91	0.0004	0.89	0.71, 1.13	0.36	0.43

Abbreviations: BMI, body mass index; CI, confidence interval; HR, hazard ratio; MET, metabolic equivalent; PMH, postmenopausal hormone.

^a The total of colon cancers and rectal cancers does not equal the total of colorectal cancers due to some cases lacking a verified location.

^b For variables whose β coefficients were calculated per year, the relative risk accounts for 40 years (e.g., ages 30–70 years) and for the specified contrast. Relative risk = $\exp[\beta (40 \text{ years}) (\text{contrast})]$. For dichotomous variables, the relative risk was calculated as $\exp(\beta)$.

^c Family history of colon or rectal cancer in a first-degree relative.

^d BMI was calculated as weight (kg)/height (m)².

^e Sigmoidoscopy or colonoscopy.

was statistically heterogeneous for colon cancer and rectal cancer; the hazard ratio for colon cancer (HR = 0.82) was significantly higher than the hazard ratio for rectal cancer (HR = 0.48; $P < 0.0001$ and P for colon vs. rectal = 0.0008).

Table 3 shows results by more finely defined subsites within the colon (e.g., proximal and distal colon) and the P values for pair-wise comparisons between the sites. Several pair-wise tests for heterogeneity for age were significant; however, all subsite hazard ratios were statistically significant and in the same direction (HR for proximal = 2.01, HR for distal = 1.62, and HR for rectal = 1.71). Red meat (unprocessed) had a direct association with proximal colon cancer (HR = 1.15) and an inverse association with distal colon cancer (HR = 0.82) that resulted in a borderline significant test for heterogeneity (P for proximal vs. distal = 0.068). Processed meat, which did not show an overall association (Table 2), revealed a statistically significant association with distal colon cancer (HR = 1.45; $P = 0.02$) but not proximal colon cancer (HR = 0.95; $P = 0.72$ and P for proximal vs. distal = 0.050) (Figure 1).

We observed statistically significant heterogeneity for smoking by subsite within the colon: 40 cumulative pack-years was associated with proximal colon cancer (HR = 1.31; $P < 0.0001$) but not distal colon cancer (HR = 1.04; $P = 0.66$ and P for proximal vs. distal = 0.029). Figure 2 displays the 20-year cumulative incidence of proximal and distal colon cancer by smoking status and shows the site-specific associations.

Physical activity showed heterogeneity by subsite (P for distal vs. rectal = 0.017), with a more inverse association with distal colon cancer (HR = 0.40) than with rectal cancer (HR = 0.89). Screening coverage was not associated with proximal colon cancer, but it was inversely associated with distal colon and rectal cancer (HR for proximal = 0.96, HR for distal = 0.55, and HR for rectal = 0.48; $P < 0.0001$ for both proximal vs. distal and proximal vs. rectal). Calcium intake was inversely associated with distal colon cancer only (HR = 0.65; $P < 0.0001$ and P for distal vs. rectal = 0.049).

Table 3. Hazard Ratios for Incident Proximal and Distal Colon Cancer and Rectal Cancer (Competing-Risks Model) in Nurses' Health Study, 1980–2010

Risk Factor	Proximal Colon (n = 821)			Distal Colon (n = 521)			Rectal (n = 376)			P Value		
	HR ^a	95% CI	P Value	HR ^a	95% CI	P Value	HR ^a	95% CI	P Value	Proximal vs. Distal	Distal vs. Rectal	Proximal vs. Rectal
Age (years; 60 vs. 50)	2.01	1.83, 2.21	<0.0001	1.62	1.45, 1.81	<0.0001	1.71	1.50, 1.96	<0.0001	0.004	0.52	0.053
Family history of colon or rectal cancer (yes vs. no) ^b	1.55	1.31, 1.83	<0.0001	1.43	1.15, 1.79	0.002	1.26	0.96, 1.66	0.095	0.59	0.48	0.21
Red meat intake (servings/day per year; 1 vs. 0)	1.15	0.93, 1.42	0.19	0.82	0.61, 1.10	0.19	0.99	0.70, 1.40	0.96	0.068	0.089	0.47
Processed meat intake (servings/day per year; 1 vs. 0)	0.95	0.72, 1.26	0.72	1.45	1.06, 1.98	0.02	1.13	0.75, 1.71	0.57	0.050	0.35	0.50
Folate intake (µg/day per year; 600 vs. 200)	0.80	0.67, 0.96	0.016	0.92	0.74, 1.16	0.49	0.77	0.58, 1.02	0.071	0.35	0.095	0.79
Smoking history (total pack-years; 40 vs. 0)	1.31	1.16, 1.48	<0.0001	1.04	0.88, 1.23	0.66	1.27	1.05, 1.53	0.013	0.029	0.12	0.76
BMI (units per year; 30 vs. 20) ^c	1.35	1.10, 1.65	0.004	1.28	0.98, 1.67	0.067	1.50	1.10, 2.04	0.011	0.77	0.46	0.58
Physical activity level (MET-hours/week per year; 21 vs. 2)	0.64	0.47, 0.87	0.004	0.40	0.25, 0.64	0.0002	0.89	0.56, 1.41	0.63	0.11	0.017	0.23
Height (inches per year; 67 vs. 61)	1.24	1.04, 1.49	0.019	1.25	0.98, 1.60	0.070	1.24	0.93, 1.65	0.15	0.95	0.94	0.98
Alcohol intake (g/day per year; 30 vs. 0)	1.09	0.88, 1.36	0.43	1.05	0.77, 1.43	0.75	1.33	0.96, 1.85	0.088	0.84	0.30	0.33
Aspirin use (tablets per week per year; 7 vs. 0)	0.75	0.64, 0.87	0.0002	0.79	0.65, 0.97	0.023	0.80	0.63, 1.01	0.060	0.67	0.96	0.66
Endoscopic screening (yes vs. no; 20 years vs. 0)	0.96	0.84, 1.09	0.53	0.55	0.43, 0.69	<0.0001	0.48	0.36, 0.64	<0.0001	<0.0001	0.49	<0.0001
PMH use (vs. never users)												
Current use	0.95	0.79, 1.14	0.59	0.79	0.63, 1.01	0.058	0.88	0.68, 1.15	0.35	0.24	0.56	0.65
Past use	0.98	0.83, 1.16	0.82	1.03	0.84, 1.28	0.75	0.79	0.61, 1.02	0.074	0.70	0.11	0.16
Calcium intake (mg/day per year; 1,000 vs. 500)	0.90	0.77, 1.04	0.14	0.65	0.53, 0.80	<0.0001	0.89	0.71, 1.13	0.36	0.15	0.049	0.99

Abbreviations: BMI, body mass index; CI, confidence interval; HR, hazard ratio; MET, metabolic equivalent; PMH, postmenopausal hormone.

^a For variables whose β coefficients were calculated per year, the relative risk accounts for 40 years (e.g., ages 30–70 years) and for the specified contrast. Relative risk = $\exp[\beta (40 \text{ years}) (\text{contrast})]$. For dichotomous variables, the relative risk was calculated as $\exp(\beta)$.

^b Family history of colon or rectal cancer in a first-degree relative.

^c BMI was calculated as weight (kg)/height (m)².

Table 4 shows the results obtained when we evaluated the cecum separately and conducted 2 trend tests: one for the cecum to the distal colon (trend 1) and the second for the cecum to the rectum (trend 2). Generally, the estimates for the cecum were similar to those for the rest of the proximal colon (column “P Value: Cecal vs. Proximal”). Three risk factors—smoking status ($P = 0.040$ for trend 1), screening coverage ($P = 0.033$ for trend 1), and calcium intake ($P = 0.009$ for trend 1)—suggested significant trends from higher to lower associations across the cecum, proximal colon, and distal colon. Current PMH use had a borderline P value for trend 1 of 0.058. When testing for trend across the entire colorectum (trend 2), only screening coverage remained statistically significant ($P = 0.0001$ for trend 2).

Table 5 shows the C statistics for CRC and by subsite. The model performed well for the 3 main subsites: colorectum ($C = 0.607$), colon ($C = 0.603$), and rectum ($C = 0.639$). The C statistic for colon cancer was similar to what we reported for our previous model ($C = 0.61$) (20). Removing screening from the model resulted in a statistically significant reduction in the C statistic for CRC ($P = 0.0002$). The C statistic for colon cancer was statistically significantly lower than that for rectal cancer in the full model ($P = 0.03$). However, after

removing the screening variable from the model, the C statistics were similar (C for colon = 0.599 and C for rectal = 0.603; $P = 0.80$). We found no significant difference between the C statistics for proximal and distal colon cancer ($P = 0.16$).

The C statistic for the <70 years age group was statistically significantly higher than that for the ≥ 70 years age group for CRC ($P = 0.013$), colon cancer ($P = 0.015$), and proximal colon cancer ($P = 0.030$), with or without inclusion of the screening variable. The C statistics for both distal and rectal cancer did not vary significantly by age group (with or without screening).

Figure 3 displays the cumulative incidence for hypothetical risk factor profiles and the association with screening in the high-risk group. The cumulative incidences in the low-, moderate-, and high-risk groups were well-differentiated. The cumulative incidence of colon cancer was slightly reduced after including screening, but it was still higher than that for the moderate-risk profile (Figure 3A). Across the subsites (Figure 3B–C), screening was increasingly associated with lowering cumulative incidence away from that of a high-risk profile toward that of a moderate-risk profile. For rectal cancer (Figure 3D), the cumulative incidence for the individual with the high-risk profile who

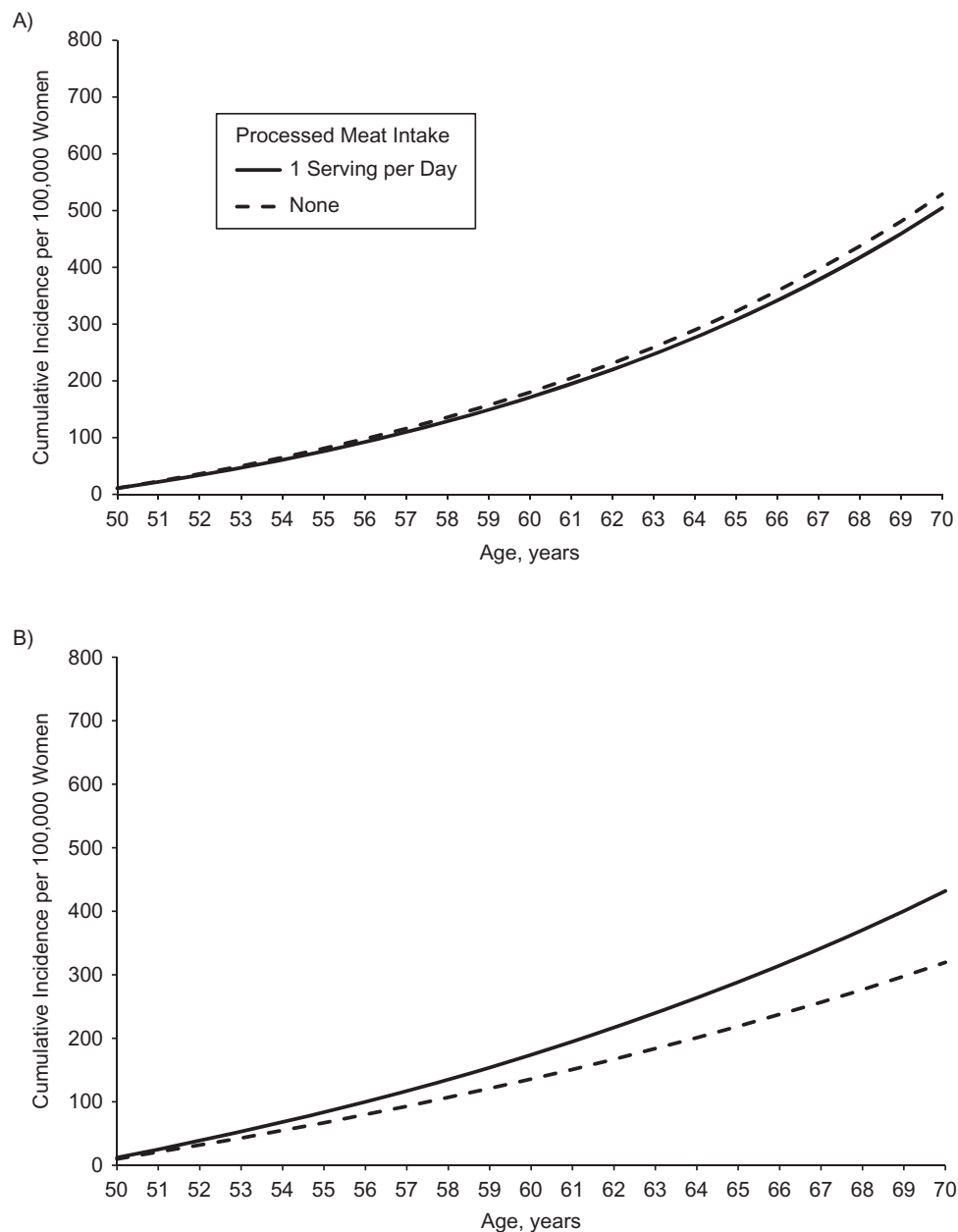


Figure 1. Cumulative incidence per 100,000 hypothetical women of proximal colon cancer (A) and distal colon cancer (B) for a woman consuming 1 serving per day of processed meat versus 0 servings per day, based on data from the Nurses' Health Study, 1980–2010. Unless otherwise specified, all women were assumed at baseline to be aged 50 years, have no family history of colon or rectal cancer in a first-degree relative, consume 1 serving per day of red meat, consume no processed meat, consume 400 $\mu\text{g/day}$ of folate, have a body mass index (weight (kg)/height (m)²) of 26, perform 15 metabolic equivalent hours/week of physical activity, have a height of 64.5 inches, consume no alcohol, use no aspirin, have had no screening endoscopy, have used no postmenopausal hormone, and consume 1,000 mg/day of calcium.

was screened was nearly equivalent to that of the individual with the moderate-risk profile.

DISCUSSION

We have updated our previous incidence model for colon cancer to include CRC, rectal cancer, longer follow-up,

additional cases (1,342 vs. 701 colon cases), modifications to our exposure variables (e.g., separate meat variables, adding alcohol intake and calcium intake, and a different smoking variable), outlier exclusion, and results by anatomical subsite within the colon, with statistical testing for heterogeneity.

Compared with our previous model, we observed 1) a lower point estimate for red and processed meat (albeit not

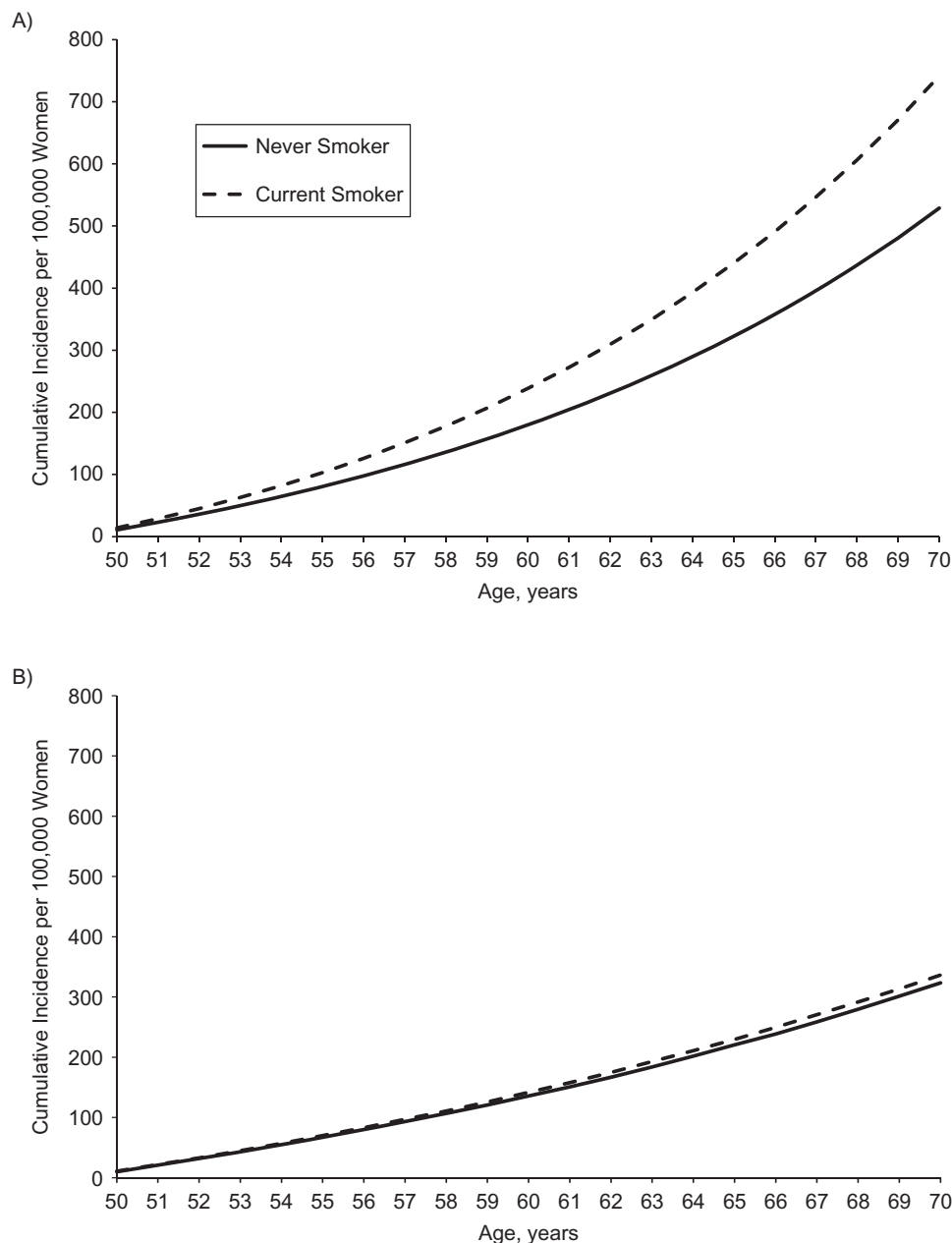


Figure 2. Cumulative incidence per 100,000 hypothetical women of proximal colon cancer (A) and distal colon cancer (B) for a never smoker versus a current smoker (30 pack-years by age 30 years and 1 pack/day from ages 30–70 years), based on data from the Nurses' Health Study, 1980–2010. Unless otherwise specified, all women were assumed at baseline to be aged 50 years, have no family history of colon or rectal cancer in a first-degree relative, consume 1 serving per day of red meat, consume no processed meat, consume 400 $\mu\text{g/day}$ of folate, have a body mass index (weight (kg)/height (m)²) of 26, perform 15 metabolic equivalent hours/week of physical activity, have a height of 64.5 inches, consume no alcohol, use no aspirin, have had no screening endoscopy, have used no postmenopausal hormone, and consume 1,000 mg/day of calcium.

significant in either model); 2) a significant association between BMI and colon cancer risk ($P = 0.09$ in previous model); 3) a significant inverse association between folate intake and colon cancer risk; and 4) a statistically significant inverse association for calcium intake (variable not included in the previous model).

Risk factors for rectal and colon cancer were generally similar. Our findings of possible heterogeneity by subsite for physical activity is consistent with a recent meta-analysis (30). Alcohol intake was not statistically significantly associated with colon or rectal cancer, although the hazard ratio for rectal cancer was suggestive of a direct association. In

Table 4. Hazard Ratios for Incident Cecal, Proximal, and Distal Colon Cancer and Rectal Cancer (Competing-Risks Model) in the Nurses' Health Study, 1980–2010

Risk Factor	Cecal (n = 247)		Proximal Colon (Not Cecal) (n = 574)		Distal Colon (n = 521)		Rectal (n = 376)		P Value				
	HR ^a	P Value	HR ^a	P Value	HR ^a	P Value	HR ^a	P Value	Cecal vs. Proximal	Cecal = Proximal = Distal ^b	Cecal = Proximal = Distal = Rectal ^c	Trend 1 ^d	Trend 2 ^e
Age (years; 60 vs. 50)	1.78	<0.0001	2.12	<0.0001	1.62	<0.0001	1.71	<0.0001	0.090	0.003	0.006	0.47	0.20
Family history of colon or rectal cancer (yes vs. no) ^f	1.53	0.007	1.56	<0.0001	1.43	0.002	1.26	0.095	0.93	0.86	0.66	0.75	0.29
Red meat intake (servings/day per year; 1 vs. 0)	0.97	0.90	1.22	0.11	0.82	0.19	0.99	0.96	0.36	0.12	0.072	0.56	0.61
Processed meat intake (servings/day per year; 1 vs. 0)	0.92	0.78	0.96	0.81	1.45	0.02	1.13	0.57	0.91	0.15	0.28	0.18	0.32
Folate intake (µg/day per year; 600 vs. 200)	0.78	0.15	0.81	0.05	0.92	0.49	0.77	0.071	0.86	0.63	0.30	0.44	0.92
Smoking history (total cumulative pack-years; 40 vs. 0)	1.41	0.003	1.27	0.001	1.04	0.66	1.27	0.013	0.46	0.069	0.14	0.040	0.25
BMI (units per year; 30 vs. 20) ^g	1.52	0.025	1.28	0.042	1.28	0.067	1.50	0.011	0.44	0.71	0.76	0.45	0.98
Physical activity level (MET-hours/week per year; 21 vs. 2)	0.65	0.14	0.63	0.014	0.40	0.0002	0.89	0.63	0.93	0.27	0.13	0.21	0.69
Height (inches per year; 67 vs. 61)	1.08	0.66	1.31	0.012	1.25	0.070	1.24	0.15	0.34	0.63	0.82	0.47	0.64
Alcohol intake (g/day per year; 30 vs. 0)	0.95	0.82	1.15	0.28	1.05	0.75	1.33	0.088	0.46	0.75	0.62	0.69	0.30
Aspirin use (tablets/week per year; 7 vs. 0)	0.79	0.094	0.73	0.0009	0.79	0.023	0.80	0.060	0.68	0.84	0.93	0.99	0.82
Endoscopic screening (yes vs. no; 20 years vs. 0)	0.82	0.14	1.01	0.85	0.55	<0.0001	0.48	<0.0001	0.16	<0.0001	<0.0001	0.033	0.0001
PMH use (vs. never users)													
Current use	1.17	0.35	0.87	0.22	0.79	0.058	0.88	0.35	0.14	0.16	0.30	0.058	0.17
Past use	1.04	0.79	0.96	0.67	1.03	0.75	0.79	0.074	0.65	0.84	0.40	0.95	0.23
Calcium intake (mg/day per year; 1,000 vs. 500)	1.02	0.88	0.85	0.066	0.65	<0.0001	0.89	0.36	0.25	0.026	0.049	0.009	0.24

Abbreviations: BMI, body mass index; HR, hazard ratio; MET, metabolic equivalent; PMH, postmenopausal hormone.

^a For variables whose β coefficients were calculated per year, the relative risk accounts for 40 years (e.g., ages 30–70 years) and for the specified contrast. Relative risk = $\exp[\beta (40 \text{ years}) (\text{contrast})]$. For dichotomous variables, the relative risk was calculated as: $\exp(\beta)$.

^b β for cecum = β for proximal = β for distal.

^c β for cecum = β for proximal = β for distal = β for rectal.

^d Trend 1: *P* value for linear trend of subsite-specific beta coefficients—colon only. Ordinal distance values from anus for colon only: 1 = sigmoid, 7 = cecum. Distal colon (sigmoid and descending colon), average distance = 1.5; proximal colon (splenic flexure, transverse colon, hepatic flexure, ascending colon), average distance = 4.5. Average distance for distal, proximal, and cecal segments = $(1.5 + 4.5 + 7) \div 3 = 4.33$. Example of TEST statement for age: $(7 - 4.33) \beta_0 \text{ age_ce} + (4.5 - 4.33) \beta_1 \text{ age_pr} + (1.5 - 4.33) \beta_2 \text{ age_d} = 0$; *P* = 0.47.

^e Trend 2: *P* value for linear trend of subsite-specific beta coefficients—entire colorectum. Ordinal distance values from anus for entire colorectum: 1.5 = rectum/rectosigmoid, 9 = cecum. Distal colon (sigmoid and descending colon), average distance = 3.5; proximal colon (splenic flexure, transverse colon, hepatic flexure, ascending colon), average distance = 6.5. Average distance for rectal, distal, proximal, and cecal segments = $20.5 \div 4 = 5.125$. Example of TEST statement for age: $(9 - 5.125) \beta_0 \text{ age_ce} + (6.5 - 5.125) \beta_1 \text{ age_pr} + (3.5 - 5.125) \beta_2 \text{ age_d} + (1.5 - 5.125) \beta_3 \text{ age_r} = 0$; *P* = 0.20.

^f Family history of colon or rectal cancer in a first-degree relative.

^g BMI was calculated as weight (kg)/height (m)².

Table 5. Age-Adjusted C Statistics for a Model of Colorectal Cancer Incidence, Overall and by Tumor Subsite and Age^a and With and Without the Inclusion of Screening, Nurses' Health Study, 1980–2010

Analysis and Cancer Subgroup	No. of Cases ^b	Adjusted for All Covariates ^c	P Value ^d	Adjusted for All Covariates Except Screening	P Value ^d
		C Statistic (Standard Error)		C Statistic (Standard Error)	
Colorectal ^e	1,759	0.607 (0.007)		0.599 (0.007)	0.0002 ^f
Colon	1,342	0.603 (0.008)	0.03	0.599 (0.008)	0.80
Rectal	376	0.639 (0.014)		0.603 (0.014)	
Proximal	821	0.603 (0.010)	0.16	0.602 (0.010)	0.65
Distal	521	0.625 (0.012)		0.609 (0.012)	
Subgroup and age, years					
Colorectal					
<70	1,155	0.620 (0.008)	0.013	0.610 (0.008)	0.015
≥70	604	0.584 (0.012)		0.575 (0.012)	
Colon					
<70	851	0.617 (0.010)	0.015	0.613 (0.010)	0.015
≥70	491	0.577 (0.013)		0.573 (0.013)	
Proximal					
<70	471	0.621 (0.013)	0.030	0.620 (0.013)	0.034
≥70	350	0.578 (0.015)		0.578 (0.015)	
Distal					
<70	380	0.628 (0.014)	0.68	0.618 (0.014)	0.26
≥70	141	0.617 (0.023)		0.587 (0.024)	
Rectal					
<70	277	0.640 (0.016)	0.90	0.605 (0.017)	0.83
≥70	99	0.636 (0.028)		0.598 (0.028)	

^a C statistics adjusted for age in 10-year age groups.

^b There was a total of 1,239,121 person-years among noncases, of which 1,000,230 corresponded to <70 years of age and 238,891 to ≥70 years of age.

^c All covariates included age, family history, red meat intake, processed meat intake, folate intake, calcium intake, smoking history, body mass index, physical activity level, height, alcohol intake, aspirin use, and PMH use.

^d P value for the difference between C statistics.

^e The total number of colorectal cancer cases does not equal the sum of the total numbers of colon and rectal cancer cases due to incomplete information on tumor site.

^f P value for the difference between C statistics with and without screening.

a recent review, Hjartaker et al. (30) reported that alcohol intake was more strongly associated with rectal cancer than with colon cancer across 10 articles (especially among men). Our results could be due to a relatively low and narrow range of intake and the fact that we modeled intake continuously, whereas previous studies have often evaluated high intake (e.g., >30 g/day) (31). Folate intake was significantly inversely associated with colorectal and colon cancer but not rectal cancer. The results were similar to those from a lagged analysis of total folate intake and risk of CRC (32); although the hazard ratios were similar by subsite, the P value for heterogeneity was not significant. Screening was statistically significantly associated with risk of distal colon and rectal cancer and reflected a linear association from the proximal colon to the rectum. The lack of association with cecal and proximal cancer could reflect a variety of factors, including nonspecific data on endoscopy in the early follow-up period, reduced endoscopic screening visualization of the

cecum and proximal colon, and possibly the higher incidence of serrated polyps in the proximal colon (33, 34).

The C statistic for colon cancer was similar to that in our previous model; when accounting for screening, the C statistic for rectal cancer was statistically significantly higher than that for colon cancer. Excluding screening, the model's discriminatory ability for CRC was significantly reduced. Notably, the C statistic was lower among the ≥70 years age group for colorectal, colon, and proximal colon cancer. More detailed investigation into the roles of age and screening in the discriminatory ability of prediction models is warranted.

Within the colon, smoking history had a significant association with proximal colon cancer and an even higher hazard ratio for cecal cancer versus noncecal proximal cancer. Our results suggest heterogeneity of risk for smoking history within the colon and evidence of a linear trend in risk from the cecum to the distal colon. These results support

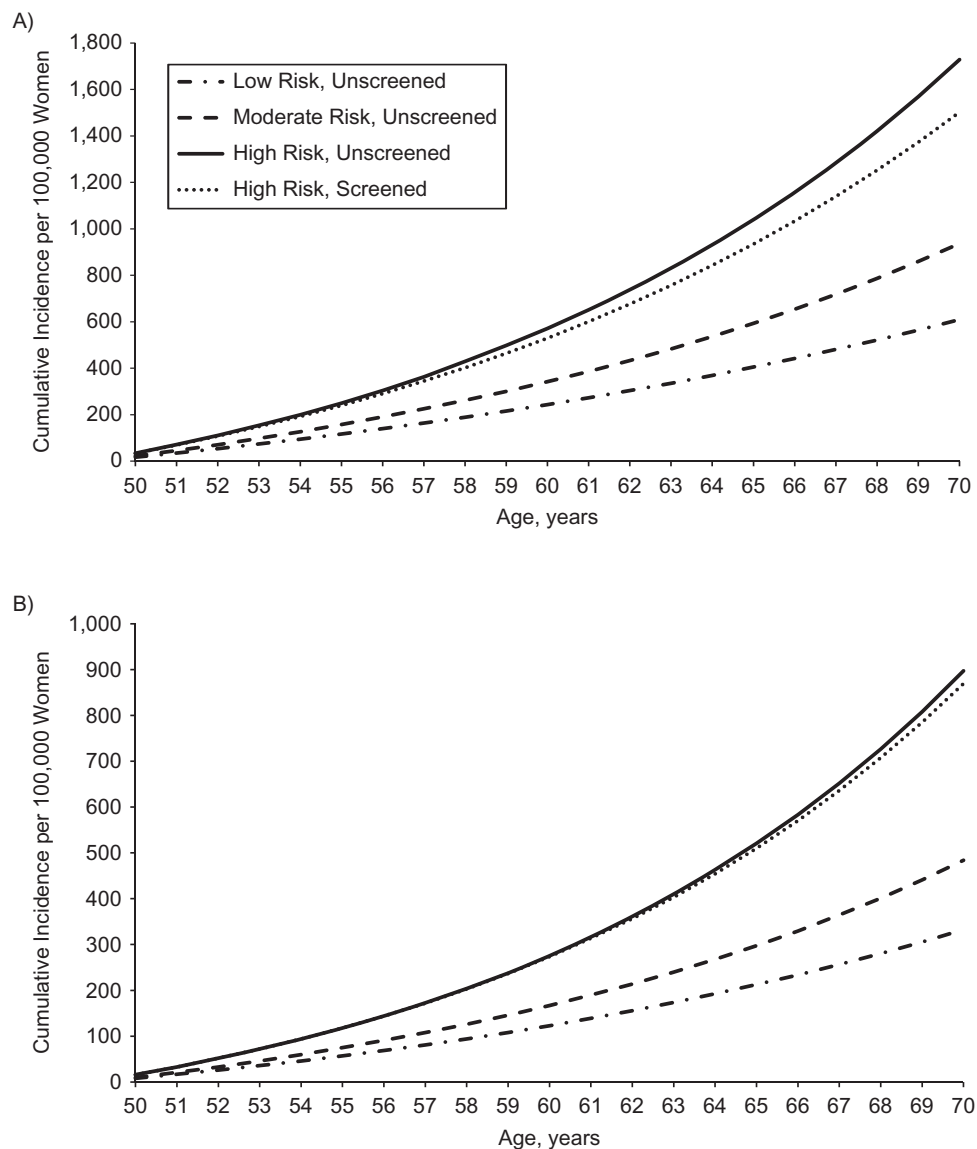


Figure 3 Continues.

the possibility that a few risk factors have a continuum of association across the colon.

We uncovered a significant association between processed meat and distal colon cancer that was in sharp contrast to the lack of association with other sites. Similar results were reported in a recent analysis on meat intake that included the Nurses' Health Study (35). In a recent review, Hjartåker et al. (30) similarly reported that red and processed meat consumption tended to be more strongly associated with the risks of distal colon and rectal cancer than with the risk of proximal cancer.

In this paper, we have graphically presented data on cumulative incidence for hypothetical unscreened low-, moderate-, and high-risk individuals, as well as on the role of screening in the high-risk group. Our results suggested that,

particularly for proximal colon cancer, lifestyle factors have a stronger association with cumulative incidence when compared with screening. In contrast, screening had a stronger association with the cumulative incidence of rectal cancer—comparable in magnitude to changing from a high-risk profile to a moderate-risk profile.

Because we evaluated over 10 risk factors and several sub-sites, we cannot exclude the possibility that some of the statistically significant associations we observed resulted from multiple comparisons. However, our overall results were generally in agreement with those of previous studies that have evaluated individual risk factors separately and that specified confounding variables in various ways. Moreover, our model is unique in its ability to update all exposure variables multiple times across the entire follow-up period.

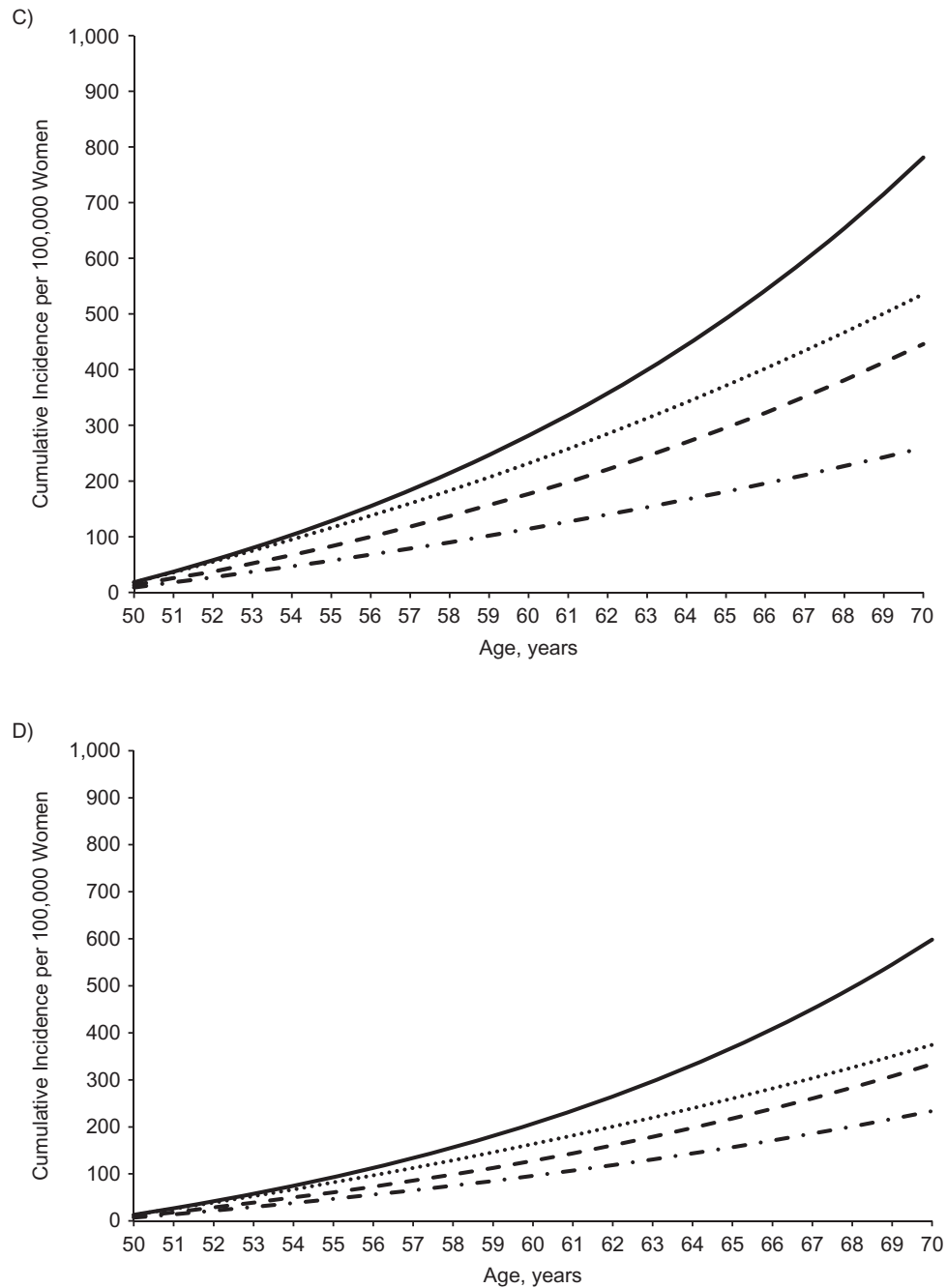


Figure 3. Cumulative incidence per 100,000 hypothetical women of total colon cancer (A), proximal colon cancer (B), distal colon cancer (C), and rectal cancer (D) for an unscreened individual with a low-risk profile, moderate-risk profile, or high-risk profile and for a high-risk profile with 20 years of screening coverage, based on data from the Nurses' Health Study, 1980–2010. Unless otherwise specified, all women were assumed at baseline to be aged 50 years, have a height of 64.5 inches, consume no alcohol, use no aspirin, have had no screening endoscopy, and have used no postmenopausal hormone. Low-risk profile: never smoker, body mass index (BMI) (weight (kg)/height (m)²) of 21, physical activity level of 21 metabolic equivalent (MET) hours/week, 0 servings per day of red and processed meat, folate intake of 600 μ g/day, and calcium intake of 1,000 mg/day. Moderate-risk profile: never smoker, BMI of 26, physical activity level of 15 MET-hours/week, 0 servings per day of red and processed meat, folate intake of 400 μ g/day, and calcium intake of 750 mg/day. High-risk profile: 15 pack-years accumulated by age 30 and 1 pack per day from ages 50–70 years, BMI of 30, physical activity level of 2 MET-hours/week, 1 serving of red meat and 1 serving of processed meat per day, folate intake of 200 μ g/day, and calcium intake of 500 mg/day.

Although anatomical definitions have been used in clinical, pathological, and epidemiologic settings (15, 16), newer evidence supports a continuum of risk associations throughout the colorectum and the importance of molecular characteristics (13, 14). The prevalences of CpG island methylator phenotype–high, microsatellite instability–high, and *BRAF* and *PIK3CA* mutations have been shown to increase linearly along the colorectal subsites from rectum to ascending colon, and cecal cancer have the highest prevalence of *KRAS* mutations (14). Results from other studies on CRC subtypes and tumor molecular features have been generally consistent (18, 36–38), attesting to the biological heterogeneity between CRCs at different subsites. Risk prediction models for various molecular subtypes of CRC may provide additional etiological insight (39).

Our results support an overall similar risk factor profile for colon and rectal cancer. Although rectal cancer has been studied less widely than colon cancer, the risk factors appear to largely overlap; thus, recommendations for colorectal or colon cancer prevention should largely apply to rectal cancer. The possibility of variation for specific risk factors (e.g., smoking history and meat intake) merits further evaluation. Although screening is known to reduce CRC incidence and mortality (40), we observed a substantial role of lifestyle changes in the cumulative incidence of CRC; promoting healthy lifestyles should remain a priority for CRC prevention. Lastly, although our model overall had reasonable discriminatory ability throughout the colorectum, inclusion of novel risk factors may improve its predictive ability. Given the increasing evidence that early-life risk factors (41–44) and timing of exposures (45, 46) influence CRC risk, this model could also eventually incorporate risk factors over the life course and also be adapted for more diverse populations (47). Consistently lower discriminatory ability (*C* statistics) in older age groups suggests a need for further study of the age-specific etiology of CRC.

An important step in the development of prediction models is external validation in an independent sample (10, 12, 36, 48, 49). In conjunction with a parallel model we have developed in men (using data from the Health Professionals Follow-up Study), our next steps are to externally validate our model in the National Institutes of Health–AARP Diet and Health Study and then develop the model for implementation in clinical settings.

ACKNOWLEDGMENTS

Author affiliations: Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts (Esther K. Wei, Edward L. Giovannucci, Robert J. Glynn, Charles S. Fuchs, Meir Stampfer, Walter Willett, Bernard Rosner); California Pacific Medical Center Research Institute, San Francisco, California (Esther K. Wei); Alvin J. Siteman Cancer Center at Barnes-Jewish Hospital and Washington University School of Medicine, St. Louis, Missouri (Graham A. Colditz); Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, Massachusetts (Edward L. Giovannucci, Kana Wu, Meir Stampfer, Walter Willett); Department of

Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts (Edward L. Giovannucci, Meir Stampfer, Walter Willett, Shuji Ogino); Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts (Robert J. Glynn, Bernard Rosner); Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, Massachusetts (Charles S. Fuchs, Shuji Ogino); and Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts (Shuji Ogino).

This work was supported by the National Cancer Institute at the National Institutes of Health (grants UM1 CA186107, P01 CA87969, R21 CA158808, P50 CA127003, R35 CA197735, R01 CA151993, and P30 CA006516); the Paula and Russell Agrusa Fund for Colorectal Cancer Research; the Friends of Dana-Farber Cancer Institute; the Bennett Family Foundation; and the Entertainment Industry Foundation, through the National Colorectal Cancer Research Alliance.

We thank the following state cancer registries for their help: Alabama, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, Florida, Georgia, Idaho, Illinois, Indiana, Iowa, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Nebraska, New Hampshire, New Jersey, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, Tennessee, Texas, Virginia, Washington, and Wyoming.

Conflict of interest: none declared.

REFERENCES

1. Costantino JP, Gail MH, Pee D, et al. Validation studies for models projecting the risk of invasive and total breast cancer incidence. *J Natl Cancer Inst.* 1999;91(18):1541–1548.
2. Freedman AN, Seminara D, Gail MH, et al. Cancer risk prediction models: a workshop on development, evaluation, and application. *J Natl Cancer Inst.* 2005;97(10):715–723.
3. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst.* 1989;81(24):1879–1886.
4. Rosner B, Colditz GA, Iglehart JD, et al. Risk prediction models with incomplete data with application to prediction of estrogen receptor–positive breast cancer: prospective data from the Nurses' Health Study. *Breast Cancer Res.* 2008;10(4):R55.
5. Colditz GA, Rosner B. Cumulative risk of breast cancer to age 70 years according to risk factor status: data from the Nurses' Health Study. *Am J Epidemiol.* 2000;152(10):950–964.
6. Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. *Stat Med.* 2004;23(7):1111–1130.
7. Chen WY. Exogenous and endogenous hormones and breast cancer. *Best Pract Res Clin Endocrinol Metab.* 2008;22(4):573–585.
8. Hankinson SE, Colditz GA, Willett WC. Towards an integrated model for breast cancer etiology: the lifelong interplay of genes, lifestyle, and hormones. *Breast Cancer Res.* 2004;6(5):213–218.
9. Kobayashi S, Sugiura H, Ando Y, et al. Reproductive history and breast cancer risk. *Breast Cancer.* 2012;19(4):302–308.

10. Park Y, Freedman AN, Gail MH, et al. Validation of a colorectal cancer risk prediction model among white patients age 50 years and older. *J Clin Oncol*. 2009;27(5):694–698.
11. Driver JA, Gaziano JM, Gelber RP, et al. Development of a risk score for colorectal cancer in men. *Am J Med*. 2007;120(3):257–263.
12. Steffen A, MacInnis RJ, Joshy G, et al. Development and validation of a risk score predicting risk of colorectal cancer. *Cancer Epidemiol Biomarkers Prev*. 2014;23(11):2543–2552.
13. Yamauchi M, Lochhead P, Morikawa T, et al. Colorectal cancer: a tale of two sides or a continuum? *Gut*. 2012;61(6):794–797.
14. Yamauchi M, Morikawa T, Kuchiba A, et al. Assessment of colorectal cancer molecular features along bowel subsites challenges the conception of distinct dichotomy of proximal versus distal colorectum. *Gut*. 2012;61(6):847–854.
15. Bufill JA. Colorectal cancer: evidence for distinct genetic categories based on proximal or distal tumor location. *Ann Intern Med*. 1990;113(10):779–788.
16. Iacopetta B. Are there two sides to colorectal cancer? *Int J Cancer*. 2002;101(5):403–408.
17. Carethers JM. One colon lumen but two organs. *Gastroenterology*. 2011;141(2):411–412.
18. Phipps AI, Chan AT, Ogino S. Anatomic subsite of primary colorectal cancer and subsequent risk and distribution of second cancers. *Cancer*. 2013;119(17):3140–3147.
19. Jess P, Hansen IO, Gamborg M, et al. A nationwide Danish cohort study challenging the categorisation into right-sided and left-sided colon cancer. *BMJ Open*. 2013;3(5):e002608.
20. Wei EK, Colditz GA, Giovannucci EL, et al. Cumulative risk of colon cancer up to age 70 years by risk factor status using data from the Nurses' Health Study. *Am J Epidemiol*. 2009;170(7):863–872.
21. Rosner B. Percentage points for generalized ESD many-outlier procedure. *Technometrics*. 1983;25(2):165–172.
22. Feskanich D, Rimm EB, Giovannucci EL, et al. Reproducibility and validity of food intake measurements from a semiquantitative food frequency questionnaire. *J Am Diet Assoc*. 1993;93(7):790–796.
23. Troy LM, Hunter DJ, Manson JE, et al. The validity of recalled weight among younger women. *Int J Obes Relat Metab Disord*. 1995;19(8):570–572.
24. Lunn M, McNeil D. Applying Cox regression to competing risks. *Biometrics*. 1995;51(2):524–532.
25. Glynn RJ, Rosner B. Methods to evaluate risks for composite end points and their individual components. *J Clin Epidemiol*. 2004;57(2):113–122.
26. Baer HJ, Glynn RJ, Hu FB, et al. Risk factors for mortality in the Nurses' Health Study: a competing risks analysis. *Am J Epidemiol*. 2011;173(3):319–329.
27. Kang JH, Loomis SJ, Rosner BA, et al. Comparison of risk factor profiles for primary open-angle glaucoma subtypes defined by pattern of visual field loss: a prospective study. *Invest Ophthalmol Vis Sci*. 2015;56(4):2439–2448.
28. Rosner B, Glynn RJ. Power and sample size estimation for the Wilcoxon rank sum test with application to comparisons of *C* statistics from alternative prediction models. *Biometrics*. 2009;65(1):188–197.
29. Begg C, Gray R. Calculation of polychotomous logistic regression parameters using individualized regressions. *Biometrika*. 1984;71(1):11–18.
30. Hjartåker A, Aagnes B, Røsbjerg TE, et al. Subsite-specific dietary risk factors for colorectal cancer: a review of cohort studies. *J Oncol*. 2013;2013:703854.
31. Cho E, Smith-Warner SA, Ritz J, et al. Alcohol intake and colorectal cancer: a pooled analysis of 8 cohort studies. *Ann Intern Med*. 2004;140(8):603–613.
32. Lee JE, Willett WC, Fuchs CS, et al. Folate intake and risk of colorectal cancer and adenoma: modification by time. *Am J Clin Nutr*. 2011;93(4):817–825.
33. Burnett-Hartman AN, Newcomb PA, Phipps AI, et al. Colorectal endoscopy, advanced adenomas, and sessile serrated polyps: implications for proximal colon cancer. *Am J Gastroenterol*. 2012;107(8):1213–1219.
34. Payne SR, Church TR, Wandell M, et al. Endoscopic detection of proximal serrated lesions and pathologic identification of sessile serrated adenomas/polyps vary on the basis of center. *Clin Gastroenterol Hepatol*. 2014;12(7):1119–1126.
35. Bernstein AM, Song M, Zhang X, et al. Processed and unprocessed red meat and risk of colorectal cancer: analysis by tumor location and modification by time. *PLoS One*. 2015;10(8):e0135959.
36. Pfeiffer RM, Park Y, Kreimer AR, et al. Risk prediction for breast, endometrial, and ovarian cancer in white women aged 50 y or older: derivation and validation from population-based cohort studies. *PLoS Med*. 2013;10(7):e1001492.
37. Rosty C, Young JP, Walsh MD, et al. *PIK3CA* activating mutation in colorectal carcinoma: associations with molecular features and survival. *PLoS One*. 2013;8(6):e65479.
38. Rosty C, Young JP, Walsh MD, et al. Colorectal carcinomas with *KRAS* mutation are associated with distinctive morphological and molecular features. *Mod Pathol*. 2013;26(6):825–834.
39. Ogino S, Chan AT, Fuchs CS, et al. Molecular pathological epidemiology of colorectal neoplasia: an emerging transdisciplinary and interdisciplinary field. *Gut*. 2011;60(3):397–411.
40. Nishihara R, Wu K, Lochhead P, et al. Long-term colorectal-cancer incidence and mortality after lower endoscopy. *N Engl J Med*. 2013;369(12):1095–1105.
41. Nimptsch K, Giovannucci E, Willett WC, et al. Body fatness during childhood and adolescence, adult height, and risk of colorectal adenoma in women. *Cancer Prev Res (Phila)*. 2011;4(10):1710–1718.
42. Cnattingius S, Lundberg F, Iliadou A. Birth characteristics and risk of colorectal cancer: a study among Swedish twins. *Br J Cancer*. 2009;100(5):803–806.
43. Nilsen TI, Romundstad PR, Troisi R, et al. Birth size and colorectal cancer risk: a prospective population based study. *Gut*. 2005;54(12):1728–1732.
44. Sandhu MS, Luben R, Day NE, et al. Self-reported birth weight and subsequent risk of colorectal cancer. *Cancer Epidemiol Biomarkers Prev*. 2002;11(9):935–938.
45. Kim YI. Folate and colorectal cancer: an evidence-based critical review. *Mol Nutr Food Res*. 2007;51(3):267–292.
46. Wei EK, Wolin KY, Colditz GA. Time course of risk factors in cancer etiology and progression. *J Clin Oncol*. 2010;28(26):4052–4057.
47. Austin SB, Pazaris MJ, Wei EK, et al. Application of the Rosner-Wei risk-prediction model to estimate sexual orientation patterns in colon cancer risk in a prospective cohort of U.S. women. *Cancer Causes Control*. 2014;25(8):999–1006.
48. Rosner BA, Colditz GA, Hankinson SE, et al. Validation of Rosner-Colditz breast cancer incidence model using an independent data set, the California Teachers Study. *Breast Cancer Res Treat*. 2013;142(1):187–202.
49. Colditz GA, Wei EK. Risk prediction models: applications in cancer prevention. *Curr Epidemiol Rep*. 2015;2(4):245–250.