

Validation of a Colorectal Cancer Risk Prediction Model Among White Patients Age 50 Years and Older

Yikyung Park, Andrew Nathan Freedman, Mitchell H. Gail, David Pee, Albert Hollenbeck, Arthur Schatzkin, and Ruth M. Pfeiffer

ABSTRACT

Purpose

Validation of an absolute risk prediction model for colorectal cancer (CRC) by using a large, population-based cohort.

Patients and Methods

The National Institutes of Health (NIH) –American Association of Retired Persons (AARP) diet and health study, a prospective cohort study, was used to validate the model. Men and women age 50 to 71 years at baseline answered self-administered questionnaires that asked about demographic characteristics, diet, lifestyle, and medical histories. We compared expected numbers of CRC patient cases predicted by the model to the observed numbers of CRC patient cases identified in the NIH-AARP study overall and in subgroups defined by risk factor combinations. The discriminatory power was measured by the area under the receiver-operating characteristic curve (AUC).

Results

During an average of 6.9 years of follow-up, we identified 2,092 and 832 incident CRC patient cases in men and women, respectively. The overall expected/observed ratio was 0.99 (95% CI, 0.95 to 1.04) in men and 1.05 (95% CI, 0.98 to 1.11) in women. Agreement between the expected and the observed number of cases was good in most risk factor categories, except for in subgroups defined by CRC screening and polyp history. This discrepancy may be caused by differences in the question on screening and polyp history between two studies. The AUC was 0.61 (95% CI, 0.60 to 0.62) for men and 0.61 (95% CI, 0.59 to 0.62) for women, which was similar to other risk prediction models.

Conclusion

The absolute risk model for CRC was well calibrated in a large prospective cohort study. This prediction model, which estimates an individual's risk of CRC given age and risk factors, may be a useful tool for physicians, researchers, and policy makers.

J Clin Oncol 27:694-698. © 2008 by American Society of Clinical Oncology

INTRODUCTION

Freedman et al^{1a} developed a colorectal cancer (CRC) risk prediction model that estimates the probability of developing CRC given a specific age, risk factor profile, and time period (eg, 10 years) in white men and women age 50 years and older. Relative risks were estimated separately for proximal, distal, and rectal cancer by using data from population-based, case-control studies; baseline age-specific hazard rates were estimated from attributable risks from the population-based, case-control studies and from competing hazard rates from the Surveillance Epidemiology and End Results (SEER) program. The probability of developing CRC was estimated by combining competing and relative risks and baseline hazards. The projected CRC absolute risk estimate is based on an individual's age, sex, history of colorectal cancer, sigmoidoscopy/colonos-

copy, polyps, family history of CRC, smoking, physical activity, aspirin/nonsteroidal anti-inflammatory drug (NSAID) use, vegetable intake, body mass index, and hormone replacement therapy use in women. Before this prediction model can be recommended as a useful tool, it needs to be validated in an independent population. Therefore, we evaluated the performance of the CRC absolute risk prediction model in men and women in a large prospective cohort study conducted in the United States.

PATIENTS AND METHODS

Study Population

The National Institutes of Health –American Association of Retired Persons (AARP) diet and health study has been described previously.^{1,2} Briefly, the NIH-AARP study included 567,169 men and women who were 50 to 71 years old and who were residing in one of six U.S.

From the Division of Cancer Epidemiology and Genetics; and Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD; Information Management Systems, Rockville, MD; and AARP, Washington, DC.

Submitted April 2, 2008; accepted August 6, 2008; published online ahead of print at www.jco.org on December 29, 2008.

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Corresponding author: Ruth M. Pfeiffer, PhD, Biostatistics Branch, Division of Cancer Epidemiology and Genetics, 6120 Executive Blvd, Bethesda, MD 20852; e-mail: pfeiffer@mail.nih.gov.

© 2008 by American Society of Clinical Oncology

0732-183X/09/2705-694/\$20.00

DOI: 10.1200/JCO.2008.17.4813

states (California, Florida, Louisiana, New Jersey, North Carolina, and Pennsylvania) and two metropolitan areas (Atlanta, Georgia, and Detroit, Michigan) in 1995 to 1996. The participants returned a detailed, self-administered baseline questionnaire on diet, medical history, and lifestyle factors. Within 6 months from the mailing of the baseline questionnaire, we sent a second questionnaire about information on CRC screening, medication use, family history of cancer, and hormone replacement therapy in women to participants who did not have breast, colorectal, or prostate cancer at baseline and who still lived in the study areas. We based our validation on the 334,910 participants who returned both questionnaires. We additionally excluded individuals who indicated they were proxies for the intended respondents, who returned the second questionnaire 1 year after the baseline, or who had missing information on one or more of risk factors in the risk prediction model. We also limited our validation to white participants. After these exclusions, the validation cohort consisted of 155,345 men and 108,057 women. The study was approved by the National Cancer Institute (NCI) Special Studies institutional review board.

Cancer Ascertainment

Cancer patient cases during follow-up (1995 to 2003) were identified through probabilistic linkage with cancer registry databases from the original eight states and from three additional states (Arizona, Nevada, and Texas). Our case ascertainment method has been described in a previous study, which demonstrated that approximately 90% of cancer occurrences were identified through the registries.³

Incident CRC patient cases were those that had the International Classification of Disease for Oncology (3rd edition)⁴ codes C180-C184 (proximal colon), C185-C187 (distal colon), C199 and C209 (rectum), and C188-C189 and C260 (not otherwise specified).

Data Collection

At baseline, information on diet, medical history, family history of cancer, and lifestyle factors were collected through a self-administered questionnaire. More detailed information on medication use, cancer screening, and hormone replacement therapy use in women were collected in a subsequent questionnaire that was mailed within 6 months from the mailing of the baseline questionnaire. Although these two questionnaires asked many questions to comprehensively capture information related to various cancer end points, we only used the risk factors identified in the CRC risk prediction model in the validation study. CRC risk factors collected in the NIH-AARP study were comparable with those in the prediction model, except for CRC screening, history of polyps, and vigorous physical activity. The NIH-AARP study asked if an individual ever had polyps (without any time restriction) in the baseline questionnaire and inquired if an individual had CRC screening (including sigmoidoscopy, colonoscopy, or proctoscopy) during the past 3 years in the second questionnaire. On the other hand, the CRC prediction model defined CRC screening as having had a sigmoidoscopy/colonoscopy in the past 10 years and having found polyps in the past 10 years. For this validation study, we used the NIH-AARP study variables in validation of the prediction model, even though they do not measure precisely the same information. The definition of vigorous physical activity in the NIH-AARP study questionnaire was the same as in the CRC prediction model questionnaire, but the categorization of the responses differed. In the NIH-AARP study, physical activity categories were never, rarely, one to three times per month, and one to two, three to four, and five or more times per week. Thus, we modified the physical activity categories in the NIH-AARP study to fit the corresponding categories in the prediction model—less than three times per month, to 0 hours per week; one to four times per week, to either greater than 0 to 2 or greater than 2 to 4 hours per week; and at least five times per week, to greater than 4 hours per week—by randomly assigning individuals on the basis of the distribution of physical activity in the controls in the prediction model data set. We also assessed the sensitivity of our results for this categorization of physical activity and for the categorization of the screening and history of polyps.

Regular use of aspirin/NSAIDs was defined as the use of those medications at least three times per week, and estrogen-positive status in women was based on the combination of menopausal status and use of hormone replacement therapy. Vegetable intake was assessed by using a self-administered food frequency questionnaire at baseline, which was an early version of the Diet

History Questionnaire developed by the National Cancer Institute.⁵ The food frequency questionnaire asked the frequency and portion size of foods consumed during the past 12 months.

Statistical Analyses

We compared the expected (E) and the observed (O) numbers of CRC patient cases overall and in subgroups defined by age and risk factor combinations. The expected number of patient cases was calculated by summing the estimated individual absolute risk for each person predicted by the Freedman et al model, given the baseline covariate values for each person during the time from entry into the cohort to December, 31, 2003. The 95% CIs for the E/O ratio were calculated by using the normal approximation to the Poisson distributions:

$$\frac{E}{O} \cdot \exp\left(\pm 1.96 \sqrt{\frac{1}{O}}\right) \quad (1)$$

If the E/O ratio was greater than 1, the risk prediction model overestimated the risk of colorectal cancer, whereas if the E/O ratio was less than 1, the risk prediction model underestimated the risk of colorectal cancer.

We evaluated the discriminatory accuracy of the prediction model by using the area under the receiver-operating characteristic curve (AUC), also known as the concordance statistic. The value of the AUC corresponds to the probability that a randomly selected patient case has a higher predicted risk than a randomly selected control participant. We estimated the AUC values from the AARP data separately for men and women and used the nonparametric estimator in Wieand et al,⁶ which accounts for ties and provides estimates of SEs.

RESULTS

Table 1 lists the prevalence of the risk factors used in the Freedman et al model for the NIH-AARP study. The mean (10th to 90th percentile) age of participants was 63 years (55 to 70 years) in both men and women. During an average of 6.9 years of follow-up, 2,092 incident CRC patient cases were identified in men, of which 832 were in proximal colon, 679 in distal colon, 570 in rectum, and 11 not specified. In women, a total of 965 incident CRC patient cases (461 in proximal, 267 in distal, 231 in rectum, and 6 not specified) were identified.

The distribution of risk factors differed somewhat by sex. The prevalence of CRC screening in the past 3 years was 55% in men and 34% in women. When history of polyps was considered, 43% of men and 27% of women had CRC screening and never had polyps, whereas 12% of men and 7% of women had CRC screening and ever had polyps. Approximately 10% of men and women had a family history of CRC. Among individuals with family history of CRC, 21% of men and 14% of women reported screening and polyps. On the other hand, among individuals with no family history of CRC, 11% of men and 6% of women had screening and ever had polyps.

The overall E/O ratio was 0.99 (95% CI, 0.95 to 1.04) in men and 1.05 (95% CI, 0.98 to 1.11) in women (Table 2). There was little variation by age group, although there was a slight overprediction in men and women age 55 to 59 years. Agreement between the expected and the observed number of patient cases was good in most categories of risk factors, except for screening and polyp history and family history of CRC.

The CRC risk was significantly underestimated in men and women who had CRC screening and never had polyps, but it was overestimated in those who had CRC screening and ever had polyps.

Table 1. Characteristics of Participants in the NIH-AARP Diet and Health Study

Characteristic	Men		Women	
	No.	%	No.	%
No. of total participants	155,345		108,057	
Mean follow-up, years	6.9		6.9	
Age, years				
Median	63		63	
Range, 10th-90th percentiles	55-70		55-70	
< 55	12,127	8	10,105	9
55-59	30,570	20	22,776	21
60-64	41,074	26	29,436	27
≥ 65	71,574	46	45,740	42
Colorectal cancer locations				
Proximal colon	832		461	
Distal colon	679		267	
Rectum	570		231	
Not specified	11		6	
Screening during the past 3 years and polyp history				
Screening and never had polyps	66,301	43	29,321	27
No screening	69,946	45	71,150	66
Screening and ever had polyps	19,098	12	7,586	7
No. of relatives with colorectal cancer				
0	141,093	91	96,726	90
1	13,119	8	10,392	10
> 2	1,143	1	939	1
Physical activity, h/wk				
> 4	34,925	22	18,462	17
> 2 and ≤ 4	22,594	15	41,793	37
> 0 and ≤ 2	56,675	36	9,449	9
0	41,151	26	38,353	35
Aspirin/NSAID use				
None	80,917	52	64,012	59
Regular	74,428	48	44,045	41
Smoking, cigarettes/d				
0	44,788	29	48,523	45
1-10	21,242	14	19,934	18
11-19	34,589	22	20,372	19
≥ 20	54,726	35	19,228	18
Vegetable intake, servings/d				
< 5	140,973	91	95,525	88
≥ 5	14,372	9	12,532	12
Body mass index, kg/m ²				
< 25	47,970	31	50,264	47
≥ 25 and < 30	76,165	49	34,387	32
≥ 30	31,210	20	23,406	22
Estrogen status				
Negative		—	52,346	48
Positive		—	55,711	52

Abbreviations: NIH, National Institutes of Health; NSAID, nonsteroidal anti-inflammatory drug.

Risk was overestimated slightly in those who did not have CRC screening. Risk was also significantly overestimated in men and women with a family history of CRC (Table 2). In women, CRC risk was overestimated among those who were physically active, those who did not regularly use aspirin/NSAIDs, and those who had estrogen-negative status. In sensitivity analyses performed for screening, history of polyps, and the physical activity variables, results did not change appreciably (data not shown).

The discriminatory power measured by the AUC was 0.61 (95% CI, 0.60 to 0.62) in men and 0.61 (95% CI, 0.59 to 0.62) in women (Fig

1). Thus, patient cases had higher predicted risks than control participants approximately 60% of the time, overall.

DISCUSSION

We evaluated the calibration and discriminatory power of a CRC risk prediction model (ie, Freedman et al^{1a} model) by using data from a large prospective cohort study. The Freedman et al 1a model was well calibrated overall in both men and women and in most categories of

Colorectal Cancer Risk Prediction Cohort Study

Table 2. Colorectal Cancer Risk Prediction During the Follow-Up of Men and Women in the NIH-AARP Diet and Health Study

Characteristic	Risk Prediction and Analysis by Sex							
	Men				Women			
	Observed Risk	Expected Risk	E/O Ratio	95% CI	Observed Risk	Expected Risk	E/O Ratio	95% CI
Overall	2,092	2,074.9	0.99	0.95 to 1.04	965	1,008.5	1.05	0.98 to 1.11
Age, years								
< 55	75	74.7	1.00	0.79 to 1.25	37	39.4	1.06	0.77 to 1.47
55-59	240	267.4	1.11	0.98 to 1.26	102	134.6	1.32	1.09 to 1.60
60-64	503	517.8	1.03	0.94 to 1.12	259	254.5	0.98	0.87 to 1.11
≥ 65	1,274	1,215.0	0.95	0.90 to 1.01	567	579.9	1.02	0.94 to 1.10
Screening during the past 3 years and polyp history								
Screening and never had polyps	791	529.2	0.67	0.62 to 0.72	245	131.6	0.54	0.47 to 0.61
No screening	1,092	1,247.9	1.14	1.08 to 1.21	643	759.2	1.18	1.09 to 1.28
Screening and ever had polyps	209	297.8	1.42	1.24 to 1.63	77	117.7	1.53	1.22 to 1.91
No. of relatives with colorectal cancer								
0	1,870	1,772.1	0.95	0.91 to 0.99	840	854.9	1.02	0.95 to 1.09
1	197	265.8	1.35	1.17 to 1.55	113	135.8	1.20	1.00 to 1.45
2	25	37.0	1.48	1.00 to 2.19	12	17.8	1.48	0.84 to 2.61
Physical activity, h/wk								
> 4	409	412.2	1.01	0.91 to 1.11	152	202.5	1.33	1.14 to 1.56
> 2 and ≤ 4	305	284.1	0.93	0.83 to 1.04	346	389.5	1.13	1.01 to 1.25
> 0 and ≤ 2	746	762.1	1.02	0.95 to 1.10	95	86.2	0.91	0.74 to 1.11
0	632	616.4	0.98	0.90 to 1.05	372	330.3	0.89	0.80 to 0.98
Aspirin/NSAID use								
None	1,170	1,207.4	1.03	0.97 to 1.09	591	688.2	1.16	1.07 to 1.26
Regular	922	867.5	0.94	0.88 to 1.00	374	320.2	0.86	0.77 to 0.95
Vegetable intake, servings/d								
< 5	1,913	1,921.3	1.00	0.96 to 1.05	858	901.5	1.05	0.98 to 1.12
≥ 5	179	153.6	0.86	0.74 to 0.99	107	107.0	1.00	0.83 to 1.21
Body mass index, kg/m ² *								
< 25	568	551.7	0.97	0.89 to 1.05	—	—	—	—
≥ 25 and < 30	1,061	1,029.6	0.97	0.91 to 1.03	738	756.3	1.02	0.95 to 1.10
≥ 30	463	493.6	1.07	0.97 to 1.17	227	252.1	1.11	0.98 to 1.26
Smoking, cigarettes/d								
0	494	520.2	1.05	0.96 to 1.15	—	—	—	—
1-10	242	250.4	1.03	0.91 to 1.17	—	—	—	—
11-19	514	467.5	0.91	0.83 to 0.98	—	—	—	—
≥ 20	842	836.8	0.99	0.93 to 1.06	—	—	—	—
Estrogen status								
Negative	—	—	—	—	573	640.4	1.12	1.03 to 1.21
Positive	—	—	—	—	392	368.1	0.94	0.85 to 1.04

Abbreviations: NIH, National Institutes of Health; E/O ratio, expected/observed ratio; NSAID, nonsteroidal anti-inflammatory drug.
 *Body mass index for women was categorized as < 30 and ≥ 30.

risk factors. However, the model overpredicted risk in the AARP cohort in people with a family history of CRC and in those with a history of screening with polyps, and it underestimated risk in those with a history of screening but no polyps. In addition, we found that the prediction model had a modest discriminatory accuracy, as measured by the AUC or concordance statistic.

Differences between the NIH-AARP questionnaires and those used to develop the model^{7,8} may account for some of the discrepancies in calibration. The screening time period covered in the NIH-AARP questionnaire was the past 3 years, compared with the past 10 years in the Freedman et al model. Second, there was no time restriction on history of polyps in the NIH-AARP study, whereas Freedman et al used polyps found in the past 10 years. Because a person with screening and polyps in the data used by

Freedman et al may have been screened as much as 10 years before, the model might overestimate risk in a person in the NIH-AARP cohort who was screened only 3 years before and who was treated for a polyp found then.

Another disagreement between the observed and the expected number of CRC patient cases was found in family history of CRC. In the NIH-AARP study, participants with a family history of CRC tended to get screening; thus, they also may have had polyps identified. Given that family history was correlated with screening and polyp history and that the risk prediction model overpredicted the risk of screening and had polyp history, overprediction by family history is not surprising.

In women taking aspirin/NSAIDs in the NIH-AARP cohort, the model overpredicted risk by 16%. This may reflect the weak inverse

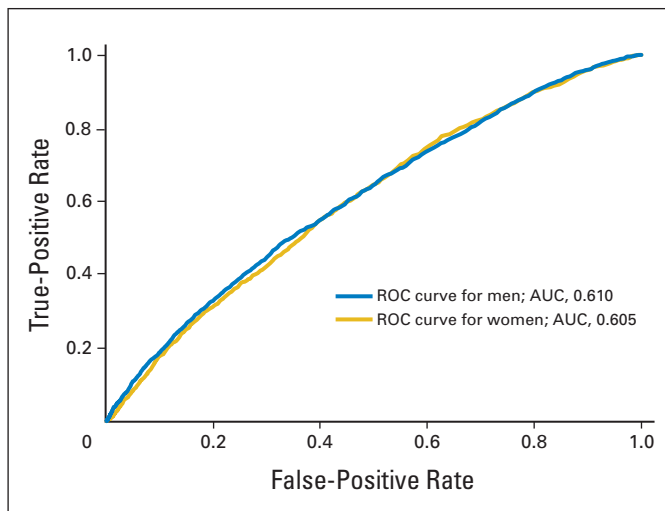


Fig 1. Receiver-operating characteristic (ROC) curves and corresponding area under the curve (AUC) statistics for the colorectal cancer absolute risk prediction model for men and women in the NIH-AARP Diet and Health Study cohort.

association between aspirin/NSAID use and CRC risk in the NIH-AARP cohort.

The discriminatory power of the CRC prediction model was modest and comparable to other cancer risk models. Studies that validate cancer risk prediction models have reported discriminatory accuracy, as measured by AUC of 0.60 to 0.63 for breast cancer,^{9,10} 0.69 for lung cancer,¹¹ 0.60 for ovarian cancer,¹² and 0.62 for melanoma.¹³ Although these cancer risk prediction models were developed on the basis of well-established risk factors for these cancers, the modest discriminatory power suggests the need to find additional strong risk predictors.

Overall, our validation results were comparable with those for other cancer risk models, such as breast and lung cancer.^{11,14} The risk prediction model was well calibrated in the NIH-AARP cohort, with some exceptions, despite some differences between this population and the population used to develop the model. The NIH-AARP study

population was younger (mostly younger than 70 years) and had a higher socioeconomic status and education level; these factors have been related to healthier lifestyle, easier access to regular cancer screening, and lower CRC risk. These differences may account for differences in the strengths of some associations with risk factors.

The relative risk features of the prediction model were based on retrospective, case-control data. It is possible that some risk factors, especially those related to behavior, may be subject to recall bias. Although the risk estimates in the prediction model were comparable to those from other published studies, the associations between risk factors and CRC tended to be stronger in the case-control studies than in the prospective cohort studies.

In conclusion, the CRC risk prediction model developed by Freedman et al was well calibrated by using a large prospective cohort study, and it had a modest discriminatory power to distinguish an individual's CRC risk. This CRC prediction model can be recommended for broader use.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The author(s) indicated no potential conflicts of interest.

AUTHOR CONTRIBUTIONS

Conception and design: Yikyung Park, Andrew Nathan Freedman, Mitchell H. Gail, Arthur Schatzkin, Ruth Pfeiffer

Administrative support: Albert Hollenbeck, Arthur Schatzkin

Provision of study materials or patients: Albert Hollenbeck, Arthur Schatzkin

Collection and assembly of data: Albert Hollenbeck, Arthur Schatzkin

Data analysis and interpretation: Yikyung Park, Andrew Nathan Freedman, Mitchell H. Gail, David Pee, Arthur Schatzkin, Ruth Pfeiffer

Manuscript writing: Yikyung Park, Andrew Nathan Freedman, Mitchell H. Gail, Ruth Pfeiffer

Final approval of manuscript: Yikyung Park, Mitchell H. Gail, Albert Hollenbeck, Arthur Schatzkin, Ruth Pfeiffer

REFERENCES

1. Freedman AN, Slattery ML, Ballard-Barbash R: A colorectal cancer risk prediction tool for white men and women without known susceptibility. *J Clin Oncol* 27:686-693, 2009
1. Schatzkin A, Subar AF, Thompson FE, et al: Design and serendipity in establishing a large cohort with wide dietary intake distributions: The National Institutes of Health-American Association of Retired Persons Diet and Health Study. *Am J Epidemiol* 154:1119-1125, 2001
2. Park Y, Subar AF, Kipnis V, et al: Fruit and vegetable intakes and risk of colorectal cancer in the NIH-AARP diet and health study. *Am J Epidemiol* 166:170-180, 2007
3. Michaud DS, Midthune D, Hermansen S, et al: Comparison of cancer registry case ascertainment with SEER estimates and self-reporting in a subset of

the NIH-AARP diet and health study. *J Registry Manag* 32:70-75, 2005

4. International Classification of Diseases for Oncology (ed 3). Geneva, Switzerland, World Health Organization, 2000

5. Diet History Questionnaire, Version 1.0. Bethesda, MD, National Institutes of Health, Applied Research Program, National Cancer Institute, 2007

6. Wieand S, Gail MH, James BR, et al: A family of nonparametric statistics for comparing diagnostic markers with paired or unpaired data. *Biometrika* 76:585-592, 1989

7. Slattery ML, Edwards SL, Ma KN, et al: Colon cancer screening, lifestyle, and risk of colon cancer. *Cancer Causes Control* 11:555-563, 2000

8. Slattery ML, Curtin KP, Edwards SL, et al: Plant foods, fiber, and rectal cancer. *Am J Clin Nutr* 79:274-281, 2004

9. Chen J, Pee D, Ayyagari R, et al: Projecting absolute invasive breast cancer risk in white women

with a model that includes mammographic density. *J Natl Cancer Inst* 98:1215-1226, 2006

10. Barlow WE, White E, Ballard-Barbash R, et al: Prospective breast cancer risk prediction model for women undergoing screening mammography. *J Natl Cancer Inst* 98:1204-1214, 2006

11. Cronin KA, Gail MH, Zou Z, et al: Validation of a model of lung cancer risk prediction among smokers. *J Natl Cancer Inst* 98:637-640, 2006

12. Rosner BA, Colditz GA, Webb PM, et al: Mathematical models of ovarian cancer incidence. *Epidemiology* 16:508-515, 2005

13. Cho E, Rosner BA, Feskanich D, et al: Risk factors and individual probabilities of melanoma for whites. *J Clin Oncol* 23:2669-2675, 2005

14. Rockhill B, Spiegelman D, Byrne C, et al: Validation of the Gail et al. model of breast cancer risk prediction and implications for chemoprevention. *J Natl Cancer Inst* 93:358-366, 2001