

# Development and Validation of a Clinical Score for Predicting Risk of Adenoma at Screening Colonoscopy

Aasma Shaukat<sup>1,2</sup>, Timothy R. Church<sup>3</sup>, Ryan Shanley<sup>4</sup>, Noah D. Kauff<sup>5</sup>, Michael J. O'Brien<sup>6</sup>, Glenn M. Mills<sup>7</sup>, Paul A. Jordan<sup>8</sup>, John A. Allen<sup>2,9</sup>, Adam Kim<sup>9</sup>, Andrew D. Feld<sup>10</sup>, Ann Graham Zauber<sup>11</sup>, and Sidney J. Winawer<sup>12</sup>

## Abstract

**Background:** Currently, no clinical tools use demographic and risk factor information to predict the risk of finding an adenoma in individuals undergoing colon cancer screening. Such a tool would be valuable for identifying those who would most benefit from screening colonoscopy.

**Methods:** We used baseline data from men and women who underwent screening colonoscopy from the randomized, multicenter National Colonoscopy Study (NCS) to develop and validate an adenoma risk model. The study, conducted at three sites in the United States (Minneapolis, MN; Seattle, WA; and Shreveport, LA) asked all participants to complete baseline questionnaires on clinical risk factors and family history. Model parameters estimated from logistic regression yielded an area under the receiver operating characteristic curve (AUROCC) used to assess prediction.

**Results:** Five hundred forty-one subjects were included in the development model, and 1,334 in the validation of the

risk score. Variables in the prediction of adenoma risk for colonoscopy screening were age (likelihood ratio test for overall contribution to model,  $P < 0.001$ ), male sex ( $P < 0.001$ ), body mass index ( $P < 0.001$ ), family history of at least one first-degree relative with colorectal cancer ( $P = 0.036$ ), and smoking history ( $P < 0.001$ ). The adjusted AUROCC of 0.67 [95% confidence interval (CI), 0.61–0.74] for the derivation cohort was not statistically significantly different from that in the validation cohort. The adjusted AUROCC for the entire cohort was 0.64 (95% CI, 0.60–0.67).

**Conclusion:** We developed and validated a simple well-calibrated risk score.

**Impact:** This tool may be useful for estimating risk of adenomas in screening eligible men and women *Cancer Epidemiol Biomarkers Prev*; 24(6): 913–20. ©2015 AACR.

## Introduction

Colorectal cancer is the second leading cause of cancer-related death in the United States (1). Current guidelines recommend initiating screening for asymptomatic men and women at age 50, using a menu of screening options (2). Most colorectal cancers are thought to arise from precursor lesions called adenomas (3–5).

<sup>1</sup>Section of Gastroenterology, VA Medical Center, Minneapolis, Minnesota. <sup>2</sup>Department of Medicine, Yale University, New Haven, Connecticut. <sup>3</sup>Division of Environmental Health Sciences, University of Minnesota, Minneapolis, Minnesota. <sup>4</sup>Division of Biostatistics, University of Minnesota, Minneapolis, Minnesota. <sup>5</sup>Ovarian Cancer Screening and Prevention, Memorial Sloan-Kettering Cancer Center, New York, New York. <sup>6</sup>Department of Pathology and Laboratory Medicine, Boston University, Boston, Massachusetts. <sup>7</sup>Section of Hematology and Oncology, Louisiana State University Health, Shreveport, Louisiana. <sup>8</sup>Department of Medicine, Louisiana State University Health, Shreveport, Louisiana. <sup>9</sup>Minnesota Gastroenterology, PA, Minneapolis, Minnesota. <sup>10</sup>Department of Gastroenterology, Group Health Cooperative, Seattle, Washington. <sup>11</sup>Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, New York. <sup>12</sup>Department of Gastroenterology and Nutrition Service, Memorial Sloan-Kettering Cancer Center, New York, New York.

**Corresponding Author:** Aasma Shaukat, University of Minnesota, 1 Veterans Drive, Minneapolis, MN 55417. Phone: 612-467-4100; Fax: 612-725-2248; E-mail: shaukat@umn.edu

doi: 10.1158/1055-9965.EPI-14-1321

©2015 American Association for Cancer Research.

The 2008 U.S. Multi-Society Task Force screening guidelines emphasized that the primary goal of screening should be prevention of colorectal cancer by detection and removal of asymptomatic adenomas (2). Recent guidelines on colorectal cancer screening by the American College of Physicians recommend that individualized risk assessment for risk of colorectal cancer should be performed in all adults, and a screening modality should be selected based on their risk (6).

Several demographic and clinical risk factors for harboring adenomas in asymptomatic men and women age 50 and over have been identified in large cohort and case-control studies and include increasing age, male sex, race, and a family history of colorectal cancer in a first-degree relative (7–10). Other identified risk factors include higher body mass index (BMI), current smoking, and heavy alcohol use (11–18). However, there is a lack of clinical tools to reliably risk-stratify men and women based on these factors. Several authors (19) have reported developing and validating risk scores for advanced neoplasia. However, there are no such tools for risk of adenomas, or that have been developed or validated in a U.S. cohort. Such clinical risk-stratification tools, or risk scores, are used not only for breast cancer (20) but also in several other areas of medicine—such as for stratifying individuals by risk of heart disease (21), for organ allocation (MELD score; ref. 22), for severity of liver disease (Child–Pugh score; ref. 23), and for hospital mortality (APACHE II; ref. 24)—where they have diagnostic or prognostic value.

Shaukat et al.

An adenoma risk score would identify the absolute risk for an individual for harboring advanced neoplasia. Based on their absolute risk, individuals could be stratified into low- and or high-risk groups, and those in the high-risk group could be prioritized for screening colonoscopy, whereas those in low-risk groups can be offered a choice of modalities of screening including colonoscopy. Given the limited capacity for colonoscopy in the United States, along with its cost and complications, the ability to risk-stratify men and women adequately would be a first step in improving resource utilization, allocating capacity, and reducing costs and complications. The objective of our study was to develop and validate a risk-prediction model by using data from a randomized multicenter clinical trial to combine the risk factors associated with adenomas into an adenoma risk score among men and women undergoing colonoscopic screening.

## Materials and Methods

We used data from phase I and II of the National Colonoscopy Study (NCS), a randomized trial of colonoscopy screening for model development and validation. The study, comparing the clinical results of colonoscopic screening compared to usual care, was conducted in two phases, between 2000 and 2002 (phase I) and between 2004 and 2007 (phase II) on a general population of men and women at three clinical centers: Group Health Cooperative, a managed care organization in the Puget Sound area of Washington State; a collaboration of the University of Minnesota and Minnesota Gastroenterology, a large group practice in Minneapolis, MN; and a wellness clinic for underserved and minorities at Louisiana State University in Shreveport, LA. The study sites were chosen to represent different areas of the country and different healthcare delivery systems. The study coordinating center was Memorial Sloan Kettering Cancer Center (MSKCC), and the central pathology center was at the Department of Pathology, Boston Medical Center (Mallory Institute of Pathology), Boston University.

Men and women age 50 years and older (40 years and older for Louisiana State University) were invited to participate. Eligibility criteria were as follows: no personal history of colorectal cancer, familial adenomatous polyposis or inflammatory bowel disease; no prior colonoscopy (phase I and II) and no prior flexible sigmoidoscopy in the previous 5 years (phase II); no serious active comorbidities such as myocardial infarction, congestive heart failure, active treatment for cancer, or on anticoagulation; no currently implanted cardioverter/defibrillator; and ability and willingness to provide informed consent. After giving consent, those eligible were randomized to colonoscopy or usual care. For the purposes of this study, we restricted our analysis to the cohort at least age 50 undergoing screening colonoscopy in phase I or II. All colonoscopies were performed by gastroenterologists with expertise in endoscopy and clinical research. Colonoscopies were performed using standard preparations and conscious sedation, with removal of all polyps. The polyps underwent histopathologic review locally and by the study pathologist (M.J. O'Brien) at the central pathology center.

Participants were asked to complete detailed questionnaires on various demographic, dietary and lifestyle factors, and on personal and family history of cancer before or shortly after their colonoscopy. The family history data were entered into a database and linked to a pedigree drawing program. Each pedigree was reviewed by a Genetics Review Committee at MSKCC to identify

participants with possible hereditary non polyposis colon cancer or familial adenomatous polyposis, who were subsequently counseled to see their physicians and discuss colorectal cancer screening.

### Statistical analysis

For the purpose of risk-score derivation and validation, only those subjects over the age of 50 at time of colonoscopy with a complete colonoscopy to the cecum and adequate bowel preparation and who completed all the relevant items on the baseline questionnaire were included.

**Risk-score derivation.** The outcome variable was one or more adenomas or colorectal cancer found by the screening colonoscopy. Because the analysis was predictive rather than causal, we restricted our analysis to individuals that had complete information for variables of interest. The independent variables were age at colonoscopy, sex, race, BMI, smoking history, current use of aspirin or NSAIDs at least once per week for a year or more, history of colorectal cancer in at least one first-degree relative, and history of colorectal cancer in at least one second-degree relative (if no first-degree relative is affected). The effects of these variables, along with interactions selected before model fitting, were estimated by logistic regression. Likelihood ratio tests, combining the main effects with the interaction effects, measured the overall effect of variables involved with interactions. In addition to the listed variables, effects of the three NCS clinical centers on the risk of adenomatous polyps were estimated in a separate regression. The receiver operating characteristic curve (ROCC) was estimated with the convex hull approach (25), and the area under this curve (AUROCC) was computed. To mitigate bias induced by such reuse of the data, the bootstrap method for estimation of prediction error was applied (26). Because the adjustments are based on bootstrap sample means of the parameter estimates, 200 iterations of bootstrap resampling were used. Confidence intervals (CI) for the adjusted estimates were produced by a second, outer bootstrap applied to the entire estimation and adjustment process, yielding a double bootstrap. To obtain adequate precision for the CIs, the outer bootstrap consisted of 2,000 iterations.

**Risk-score validation.** The risk score developed on phase I participants was applied to phase II participants. To verify the bias adjustment procedure nonparametrically, another 2,000-iteration bootstrap was applied to estimate the difference between the adjusted AUROCC from phase I and the AUROCC of the same risk score applied to phase II. This difference was not statistically significantly different from zero, establishing the adequacy of the bias correction. A new logistic regression model was then fit using combined phase I and II data, and the same double bootstrap was applied to estimate the final adjusted AUROCC and CI.

Deciles of predicted probabilities were plotted against observed relative frequencies of participants with adenomas to examine the calibration of our model. The cumulative distribution of estimated risk was plotted to determine proportions of the population falling below any particular risk.

## Results

In phase I (derivation cohort), a total of 700 men and women were enrolled at the three sites, of whom 691 filled out some part of the baseline questionnaire, and 622 underwent a colonoscopy.

**Table 1.** Phase I (derivation cohort) distribution of baseline variables

Baseline variables	Overall N (%) or mean ± SD n = 691	Individuals who underwent colonoscopy and filled some part of baseline questionnaire, by adenoma detection n = 621		Individuals included in final analysis, by adenoma detection n = 541	
		Adenoma detected n = 116	No adenoma detected n = 505	Adenoma detected n = 106	No adenoma detected n = 435
Age at baseline, y					
40–49	70 (10%)	5 (4%)	51 (10%)	0	1 (0.2%)
50–59	346 (50%)	58 (50%)	258 (51%)	54 (51%)	190 (44%)
60–69	275 (40%)	53 (46%)	196 (39%)	52 (49%)	244 (56%)
Sex					
Male	317 (46%)	66 (57%)	227 (45%)	60 (57%)	204 (47%)
Female	374 (54%)	50 (43%)	278 (55%)	46 (43%)	231 (53%)
Race					
White non-hispanic	572 (83%)	101 (88%)	421 (83%)	96 (91%)	383 (88%)
Other	116 (17%)	14 (12%)	82 (16%)	10 (9%)	52 (12%)
Missing	3 (0.5%)	1 (0.9%)	2 (0.4%)		
Center location					
Minnesota	290 (42%)	53 (46%)	223 (44%)	49 (46%)	214 (49%)
Louisiana	200 (29%)	22 (19%)	146 (29%)	17 (16%)	93 (21%)
Washington	201 (29%)	41 (35%)	136 (27%)	40 (38%)	128 (29%)
BMI					
<20	15 (2%)	2 (2%)	9 (2%)	2 (2%)	7 (2%)
20–24	198 (29%)	23 (20%)	153 (30%)	21 (20%)	143 (33%)
25–29	273 (40%)	50 (43%)	198 (39%)	47 (44%)	178 (41%)
≥30	187 (27%)	38 (33%)	130 (26%)	36 (34%)	107 (25%)
Missing	18 (3%)	3 (3%)	15 (3%)		
Colorectal cancer in at least one first-degree relative					
Yes	90 (13%)	21 (18%)	61 (12%)	20 (19%)	54 (12%)
No	599 (87%)	94 (81%)	443 (88%)	86 (81%)	381 (88%)
Missing	2 (0.3%)	1 (0.9%)	1 (0.2%)		
Colorectal cancer in at least one second-degree relative					
Yes	82 (12%)	21 (18%)	55 (11%)	20 (19%)	50 (11%)
No	607 (88%)	94 (81%)	449 (89%)	86 (81%)	385 (89%)
Missing	2 (0.3%)	1 (0.9%)	1 (0.2%)		
Smoking					
Ever	379 (55%)	69 (59%)	272 (54%)	65 (61%)	242 (56%)
Never	312 (45%)	47 (41%)	233 (46%)	41 (39%)	193 (44%)
Pack-years for ever smokers	17.2 ± 18.2	17.8 ± 20.0	17.0 ± 17.7	18.4 ± 20.4	17.5 ± 17.9
Regular intake of aspirin					
Yes	254 (37%)	41 (35%)	196 (39%)	40 (38%)	174 (40%)
No	436 (63%)	75 (65%)	308 (61%)	66 (62%)	261 (60%)
Missing	1 (0.1%)	0	1 (0.2%)		
Regular intake of NSAIDs					
Yes	175 (25%)	22 (19%)	132 (26%)	22 (21%)	113 (26%)
No	514 (74%)	94 (81%)	371 (73%)	84 (79%)	322 (74%)
Missing	2 (0.3%)	0	2 (0.4%)		

One person who underwent colonoscopy did not fill out the baseline questionnaire, so 621 individuals had both a colonoscopy and a baseline questionnaire. Of these, 541 individuals remained in the final derivation cohort after excluding those 40 to 49 years old ( $n = 56$ ) and those with missing information on their baseline questionnaire on one or more variables included in the model ( $n = 34$ ). A few individuals age 49 at randomization but 50 at colonoscopy were included.

In phase II (validation cohort), 1,763 were enrolled, of whom 1,693 were 50 years or older and 1,450 filled some part of the baseline questionnaire and underwent a colonoscopy. Of these, 1,334 had complete information on baseline variables and colonoscopy, and were included in the validation analyses. The demographic and baseline characteristics of the derivation and validation cohorts are presented in Tables 1 and 2.

Of the 541 individuals in the derivation cohort, one or more adenomas were found in 106 (19%) individuals. One or more advanced adenomatous polyps (villous histology, size larger than 1 cm, high-grade dysplasia, or cancer) were found in 33 participants (6%). Of the 1,334 individuals in the validation cohort, one or more adenomas were found in 307 (23%) individuals, and one or more advanced adenomas were found in 76 (5%) individuals. The histopathology findings at colonoscopy are presented in Table 3 for the individuals in the derivation and validation cohort included in the analyses.

The adjusted AUROCC for the phase I derivation model was 0.67 (95% CI, 0.61–0.74). That risk score applied to phase II resulted in an AUROCC of 0.61 (95% CI, 0.59–0.65). The bootstrap 95% CI for the difference of these two AUROCCs was (–0.13, +0.01), which was not significantly different from zero.

Shaukat et al.

**Table 2.** Phase II (validation cohort) distribution of baseline variables

Baseline variables	Overall N (%) or mean ± SD n = 1,609	Individuals who underwent colonoscopy and filled some part of baseline questionnaire, by adenoma detection n = 1,450		Individuals included in final analysis, by adenoma detection n = 1,334	
		Adenoma detected n = 330	No adenoma detected n = 1,120	Adenoma detected n = 307	No adenoma detected n = 1,027
Age at baseline, y					
40–49	137 (9%)	19 (6%)	94 (8%)	3 (1%)	8 (1%)
50–59	1,110 (69%)	218 (66%)	788 (70%)	214 (70%)	781 (76%)
60–69	362 (23%)	93 (28%)	283 (21%)	90 (29%)	238 (23%)
Sex					
Male	784 (49%)	199 (60%)	521 (47%)	187 (61%)	490 (48%)
Female	825 (51%)	131 (40%)	599 (53%)	120 (39%)	537 (52%)
Race					
White non-hispanic	1,301 (81%)	267 (81%)	925 (83%)	253 (82%)	884 (86%)
Other	308 (19%)	63 (19%)	195 (17%)	54 (18%)	143 (14%)
Center location					
Minnesota	954 (59%)	209 (63%)	677 (60%)	205 (67%)	675 (66%)
Louisiana	425 (26%)	87 (26%)	280 (25%)	68 (22%)	190 (19%)
Washington	230 (14%)	34 (10%)	163 (15%)	34 (11%)	162 (16%)
BMI					
<20	34 (2%)	1 (0.3%)	28 (3%)	1 (0.3%)	25 (2%)
20–24	416 (26%)	67 (20%)	314 (28%)	66 (22%)	293 (29%)
25–29	642 (40%)	141 (43%)	447 (40%)	133 (43%)	420 (41%)
≥30	513 (32%)	118 (36%)	330 (29%)	107 (35%)	289 (28%)
Missing	4 (0.3%)	3 (0.9%)	1 (0.1%)		
Colorectal cancer in at least one first-degree relative					
Yes	161 (10%)	41 (12%)	105 (9%)	37 (12%)	91 (9%)
No	1,437 (89%)	287 (87%)	1,009 (90%)	270 (88%)	936 (91%)
Missing	11 (0.7%)	2 (0.6%)	6 (0.5%)		
Colorectal cancer in at least one second-degree relative					
Yes	196 (12%)	44 (13%)	138 (12%)	37 (12%)	126 (12%)
No	1,402 (87%)	284 (86%)	976 (87%)	270 (88%)	901 (88%)
Missing	11 (0.7%)	2 (0.6%)	6 (0.5%)		
Smoking					
Ever	807 (50%)	187 (57%)	545 (49%)	174 (57%)	507 (49%)
Never	802 (50%)	143 (43%)	575 (51%)	133 (43%)	520 (51%)
Pack-years for ever smokers	18.0 ± 18.6	23.1 ± 19.7	16.3 ± 16.8	22.7 ± 19.6	16.6 ± 17.0
Regular intake of aspirin					
Yes	490 (30%)	114 (34%)	322 (29%)	106 (35%)	311 (30%)
No	1,119 (70%)	216 (66%)	798 (71%)	201 (65%)	716 (70%)
Regular intake of NSAIDs					
Yes	345 (21%)	67 (20%)	250 (22%)	60 (20%)	234 (23%)
No	1,264 (79%)	263 (80%)	870 (78%)	247 (80%)	793 (77%)

Thus, we combined phase I and II participants and re-estimated and bias-adjusted the risk score using the entire sample.

In multiple logistic regression, we found age (likelihood ratio test for overall contribution to the model,  $P < 0.001$ ), male sex ( $P < 0.001$ ), BMI ( $P < 0.001$ ), family history of at least one first-degree relative with colorectal cancer ( $P = 0.036$ ), and smoking history ( $P < 0.001$ ) to be individually associated with risk of harboring one or more adenomatous polyps in the entire cohort. The effects of clinical centers were examined in a separate regression and found to be insignificant. After expanding the model and adding interaction terms, the fitted values for all coefficients are illustrated in Fig. 1.

### Validation

The unadjusted, all-data ROCC based on our model is plotted in Fig. 2. The AUROCC was 0.64 (95% CI, 0.60–0.67). The predicted probability ranges from 0.03 to 0.7. For example, the predicted risk score for a 50-year-old female with a BMI of 20 who

is a non-smoker, has no family history of colorectal cancer, and uses aspirin daily is low (0.1), whereas that for a 61-year-old male with a BMI of 46 and 62 pack-year history, with no family history of colorectal cancer who does not use aspirin is high (0.69).

A plot of predicted risk of adenoma detection (horizontal axis) by fraction of the population at or below that risk (vertical axis) is shown in Fig. 3. This figure illustrates the impact of potential use of such a model in clinical practice: for example, if we define "high risk" as individuals where predicted probability of an adenoma is  $\geq 0.2$ , 58% would be classified as high risk and prioritized for colonoscopy, whereas 42% of individuals would be classified as low risk, and could be offered other modalities of screening. This classification would accurately capture 75% of all adenomas and 73% of advanced adenomas. Of the high-risk individuals undergoing colonoscopy, an adenoma or advanced adenoma will be found in 35%, improving the number of therapeutic screening colonoscopies, compared with all comers. Of the entire cohort, 7% of individuals harboring an adenoma or advanced adenoma would be classified as "low risk" and not be screened with

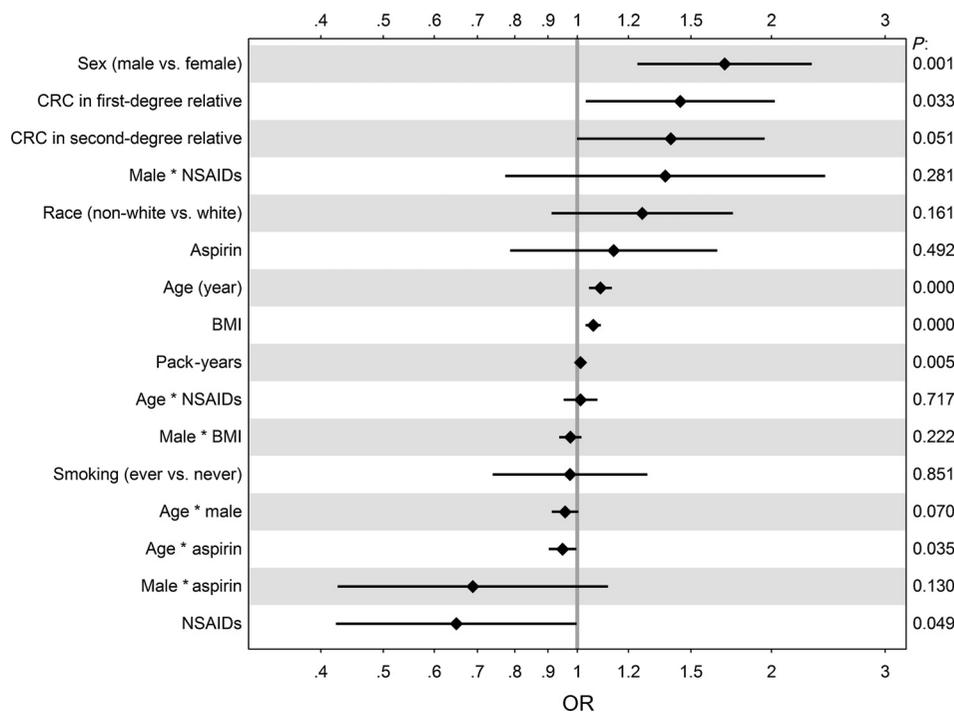
**Table 3.** Distribution of colonoscopy findings for derivation and validation cohort

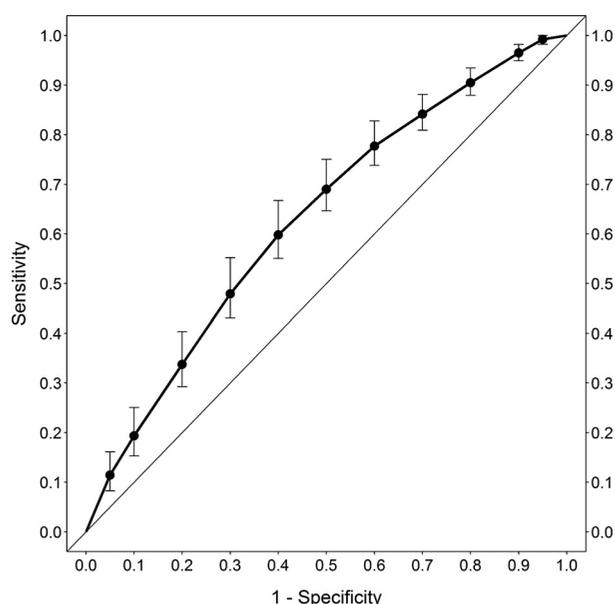
Pathology	Phase I		Phase II	
	Patients who had colonoscopy <i>n</i> = 621 <i>n</i> (%)	Patients included in analysis <i>n</i> = 541 <i>n</i> (%)	Patients who had colonoscopy <i>n</i> = 1,450 <i>n</i> (%)	Patients included in analysis <i>n</i> = 1,334 <i>n</i> (%)
Center location				
Minnesota	276 (44%)	263 (49%)	886 (61%)	880 (66%)
Louisiana	168 (27%)	110 (20%)	367 (25%)	258 (19%)
Washington	177 (29%)	168 (31%)	197 (14%)	196 (15%)
Total number of polyps per colonoscopy				
0	387 (62%)	328 (61%)	846 (58%)	781 (59%)
1	110 (18%)	91 (17%)	321 (22%)	290 (22%)
>1	124 (20%)	122 (23%)	283 (20%)	263 (20%)
Total number of adenomas per colonoscopy				
0	505 (81%)	435 (80%)	1,120 (77%)	1,027 (77%)
1	81 (13%)	72 (13%)	230 (16%)	213 (16%)
>1	35 (6%)	34 (6%)	100 (7%)	94 (7%)
Total number of right-sided adenomas per colonoscopy				
0	555 (89%)	481 (89%)	1,248 (86%)	1,146 (86%)
1	56 (9%)	50 (9%)	159 (11%)	151 (11%)
>1	10 (2%)	10 (2%)	43 (3%)	37 (3%)
Total number of left-sided adenomas per colonoscopy				
0	553 (89%)	477 (88%)	1,273 (88%)	1,169 (88%)
1	54 (9%)	51 (9%)	135 (9%)	125 (9%)
>1	14 (2%)	13 (2%)	42 (3%)	40 (3%)
Total number of advanced adenomas per colonoscopy				
0	587 (95%)	508 (94%)	1,370 (94%)	1,258 (94%)
1	30 (5%)	29 (5%)	63 (4%)	60 (4%)
>1	4 (0.6%)	4 (0.7%)	17 (1%)	16 (1%)
Total number of hyperplastic polyps per colonoscopy				
0	500 (81%)	424 (78%)	1,200 (83%)	1,104 (83%)
1	68 (11%)	65 (12%)	167 (12%)	152 (11%)
>1	53 (9%)	51 (9%)	83 (6%)	78 (6%)

colonoscopy initially. If the cutoff for "high risk" is changed to a  $\geq 0.15$  predicted probability of an adenoma, 78% would be classified as high risk and prioritized for colonoscopy, whereas 22% of

individuals would be classified as low risk. This classification would accurately capture 88% of all adenomas and 90% of advanced adenomas. Of the high-risk individuals undergoing

**Figure 1.** ORs (unadjusted) estimated from the logistic model with 95% CIs.

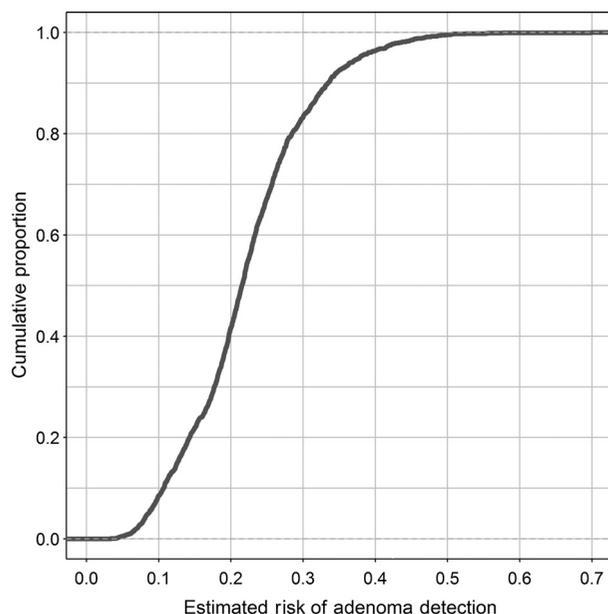




**Figure 2.**

The unadjusted, all-data ROCC based on our model Bootstrap-adjusted sensitivities is shown as black dots along with 95% CIs for given values of (1 - specificity). Our final logistic model was:  $\text{logit}[\text{Prob}(\text{adenoma})] = -1.536 + 0.057 (\text{BMI}) - 0.026 (\text{smoker}) + 0.011 (\text{pack-years}) + 0.13 (\text{aspirin}) + 0.082 (\text{age}) + 0.367 (\text{CRC FDR}) + 0.334 (\text{CRC SDR}) + 0.526 (\text{male}) - 0.432 (\text{NSAID}) + 0.232 (\text{non-white}) - 0.053 (\text{age} \times \text{aspirin}) + 0.011 (\text{age} \times \text{NSAID}) - 0.044 (\text{age} \times \text{male}) - 0.025 (\text{age} \times \text{bmi}) - 0.373 (\text{male} \times \text{aspirin}) + 0.314 (\text{male} \times \text{NSAID})$ . CRC, colorectal cancer.

colonoscopy, an adenoma or advanced adenoma will be found in 32%. Also, of the entire cohort, 3% of individuals harboring an adenoma or advanced adenoma would be classified as "low risk"



**Figure 3.**

Plot of estimated risk of adenoma detection (horizontal axis) by proportion of the population at or below that risk (vertical axis). For example, approximately 40% of our cohort had a predicted risk <0.2.

and not be screened with colonoscopy initially. As the risk criterion increases, so does the proportion offered alternatives to colonoscopy. Careful cost-benefit analyses would determine the criterion actually used.

## Discussion

Our study aimed to create and validate a risk model for quantifying an individual's risk of harboring adenomas. Our final AUROCC was 0.64, indicating that the model has good predictive utility.

Such a model could be used to determine which men and women are most likely to harbor adenomas and would likely benefit from a therapeutic colonoscopy. Although we did not directly compare colonoscopic screening with other modalities of screening, our results are a first step toward allowing payers, patients, and physicians to use risk thresholds to decide who should be offered screening colonoscopy as opposed to other modalities of screening, or who should be prioritized for screening colonoscopy, based on their estimated risk of harboring adenomas. Once externally validated in sufficient populations, and possibly improved, we envision a model such as this to be a web-based or mobile application tool readily accessible and easy to use. As illustrated in Fig. 3, a selected threshold for the cumulative probability of harboring adenomas can be used to stratify individuals for priority colonoscopy versus other modalities of screening. Combined with appropriate cost-effectiveness analysis, this information can determine the appropriate thresholds for which colonoscopic screening should be prioritized versus some other modality. For example, if we set the risk threshold for harboring an adenoma at 0.2 or above (Fig. 3), the cumulative fraction of the population below this cutoff would be about 40%, and would be offered another less invasive modality of screening, such as fecal occult blood test or even no screening. Only about 60% of the population would be prioritized to colonoscopic screening, which may greatly improve efficiency. Of course, the actual cutoff for predicted risk would have to be based on a careful analysis and comparison of the cost-effectiveness for each potential cutoff, or multiple cutoffs.

Others have used similar approaches to predict an individual's risk for harboring colorectal cancer (27). Imperiale and colleagues (28) developed a score based on three clinical factors to predict risk of harboring proximal lesions based on findings at flexible sigmoidoscopy. Others have developed risk scores for advanced neoplasia, but not adenoma, in Asian and European cohorts. Tao and colleagues (19) have used nine risk factors to derive a risk-stratification tool for a German cohort. They also took into account prior colonoscopy and detection of polyps. Kaminski and colleagues (29) used factors similar to ours: age, sex, family history of colorectal cancer, cigarette smoking, and BMI. Others have reported similar approaches in Asian patients (30–32). However, there are no clinical tools with systematic evaluation of risk factors for predicting an individual's risk of harboring an adenoma that have been performed in a U.S. cohort.

We included demographic and baseline variables that would be practical and easy to obtain for the score to be clinically useful. We also favored a parsimonious model, and included variables that others have shown to be risk factors for adenomas (8–10, 33, 34), and variables that are easy to obtain from the chart or from the patient. Also, since the risk score changes with time, it can be used to recalculate an individual's risk for harboring adenoma

periodically, and perhaps changing the individual's priority for screening colonoscopy.

Because our analysis is predictive, not causal (35), our estimated coefficients for individual factors are not necessarily unbiased estimates of their causal effects (36). Although we made no attempt to adjust for potential confounding in our model, the results were similar to findings by others, i.e., increasing age and male sex were associated with risk of adenomas (9, 27). We also found risk associations with family history of colorectal cancer in one or more first-degree relatives. In their large cohort of 3,121 mostly male veteran patients undergoing screening colonoscopy, Lieberman and colleagues (34) reported family history of colorectal cancer in one or more first-degree relatives and current smoking as risk factors for advanced adenomas (10 mm or more, 25% villous histology or more, high-grade dysplasia, or invasive cancer), whereas use of NSAIDs including aspirin was inversely associated with risk of advanced adenomas, and no association was found with BMI. Others have not found family history in a first-degree relative to be associated with adenomas (10). We did not find use of aspirin to be associated with risk of adenomatous polyps, but did find higher BMI to be associated with risk, as has been reported by others (17). We also did not find an association with non-white race and did not examine physical activity, the evidence for both of which in the literature is mixed (13, 37, 38).

The strengths of our study are the multiple sites from diverse and different geographic locations, inclusion of community-dwelling men and women in the United States, non-white races, complete colonoscopy and histology information with one central pathologist review, comprehensive collection and review of family cancer history for both first- and second-degree relatives, and community-dwelling individuals, rigorous derivation and validation of the risk score, good model calibration. The adenoma/advanced adenoma prevalence rates of 25% and 28% observed in our cohorts are similar to the rates reported by others (39, 40) and to those recommended by guidelines as indicating high quality in colonoscopy examinations (41). The high-quality colonoscopy exam, along with risk factors comparable with the general population, also supports the validity of our modeling and risk-score assessment. Our robust statistical approach included both internal and second-sample validation and a double bootstrap to address both overfitting bias and variance.

Limitations of our study include a small number of advanced adenomatous polyps precluding a separate robust model to predict only advanced adenomatous polyps and cancers. Our study is cross-sectional and predictive and thus cannot identify

causal factors for progression of adenomatous polyps to cancer. Finally, there is possibility of a recall error in self-reporting of risk factors, although this error is unlikely to be biased by outcomes because nearly all of the baseline data were collected before colonoscopy.

Although our risk score is validated and well calibrated, it would benefit from further validation in other cohorts of average risk men and women in the United States. Although the AUROCC for our model of 0.64 is good, and similar to that reported by others (29), improvements will enhance its efficiency. Development of a clinically useful risk-stratification score is important to enhancing our capacity to deliver effective colonoscopic screening targeted toward those who may benefit the most. Cost-effectiveness analyses should be performed to determine appropriate thresholds. Similar work needs to be applied to surveillance colonoscopy intervals.

### Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

### Authors' Contributions

**Conception and design:** A. Shaikat, T.R. Church, J.A. Allen, A.D. Feld, S.J. Winawer

**Development of methodology:** A. Shaikat, T.R. Church, A.G. Zauber  
**Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.):** A. Shaikat, T.R. Church, M.J. O'Brien, P.A. Jordan, J.A. Allen, A. Kim, A.D. Feld, A.G. Zauber, S.J. Winawer  
**Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis):** A. Shaikat, T.R. Church, R. Shanley, N.D. Kauff, S.J. Winawer

**Writing, review, and/or revision of the manuscript:** A. Shaikat, T.R. Church, R. Shanley, N.D. Kauff, G.M. Mills, J.A. Allen, A. Kim, A.D. Feld, A.G. Zauber, S.J. Winawer

**Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases):** A. Shaikat, T.R. Church, A.G. Zauber, S.J. Winawer

**Study supervision:** A. Shaikat, T.R. Church, S.J. Winawer

**Other (pathology review):** M.J. O'Brien

**Other (principal investigator):** A.G. Zauber

### Grant Support

This study was supported by grant R01 CA079572, National Cancer Institute, Screening Colonoscopy Feasibility Trial (National Colonoscopy Study; to A.G. Zauber and S.J. Winawer), Center for Chronic Disease Outcomes Research, and a VA HSR&D Center of Innovation (CIN 13-406; to A. Shaikat).

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 December 12, 2014; revised February 7, 2015; accepted February 9, 2015; published OnlineFirst March 23, 2015.

### References

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. *CA Cancer J Clin* 2012;62:10-29.
2. 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 Colorectal Cancer, and the American College of Radiology. *Gastroenterology* 2008;134:1570-95.
3. Winawer SJ, Zauber AG, Stewart E, O'Brien MJ. The natural history of colorectal cancer. Opportunities for intervention. *Cancer* 1991;67 (4 Suppl):1143-9.
4. Winawer SJ, Zauber AG, Ho MN, O'Brien MJ, Gottlieb LS, Sternberg SS, et al. Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. *N Engl J Med* 1993;329: 1977-81.
5. O'Brien MJ, O'Keane JC, Zauber A, Gottlieb LS, Winawer SJ. Precursors of colorectal carcinoma. Biopsy and biologic markers. *Cancer* 1992;70 (5 Suppl):1317-27.
6. Qaseem A, Denberg TD, Hopkins RH Jr, Humphrey LL, Levine J, Sweet DE, et al. Screening for colorectal cancer: a guidance statement from the American College of Physicians. *Ann Intern Med* 2012;156:378-86.

Shaukat et al.

7. Hassan C, Pickhardt PJ, Marmo R, Choi JR. Impact of lifestyle factors on colorectal polyp detection in the screening setting. *Dis Colon Rectum* 2010;53:1328–33.
8. Lieberman DA, Prindiville S, Weiss DG, Willett W, Lieberman DA, Prindiville S, et al. Risk factors for advanced colonic neoplasia and hyperplastic polyps in asymptomatic individuals. *JAMA* 2003;290:2959–67.
9. Lipkin M, Winawer SJ, Sherlock P. Early identification of individuals at increased risk for cancer of the large intestine. Part I: definition of high risk populations. *Clin Bull* 1981;11:13–21.
10. Lynch KL, Ahnen DJ, Byers T, Weiss DG, Lieberman DA, Lynch KL, et al. First-degree relatives of patients with advanced colorectal adenomas have an increased prevalence of colorectal cancer. *Clin Gastroenterol Hepatol* 2003;1:96–102.
11. Anderson JC, Stein B, Kahi CJ, Rajapakse R, Walker G, Alpern Z. Association of smoking and flat adenomas: results from an asymptomatic population screened with a high-definition colonoscope. *Gastrointest Endosc* 2010;71:1234–40.
12. Sandler RS, Lyles CM, McAuliffe C, Woosley JT, Kupper LL. Cigarette smoking, alcohol, and the risk of colorectal adenomas. *Gastroenterology* 1993;104:1445–51.
13. Hermann S, Rohrmann S, Linseisen J. Lifestyle factors, obesity and the risk of colorectal adenomas in EPIC-Heidelberg. *Cancer Causes Control* 2009;20:1397–408.
14. Anderson JC, Latreille M, Messina C, Alpern Z, Grimson R, Martin C, et al. Smokers as a high-risk group: data from a screening population. *J Clin Gastroenterol* 2009;43:747–52.
15. Kang HW, Kim D, Kim HJ, Kim CH, Kim YS, Park MJ, et al. Visceral obesity and insulin resistance as risk factors for colorectal adenoma: a cross-sectional, case-control study. *Am J Gastroenterol* 2010;105:178–87.
16. Tsilidis KK, Brancati FL, Pollak MN, Rifai N, Clipp SL, Hoffman-Bolton J, et al. Metabolic syndrome components and colorectal adenoma in the CLUE II cohort. *Cancer Causes Control*, 21:1–10.
17. Stein B, Anderson JC, Rajapakse R, Alpern ZA, Messina CR, Walker G. Body mass index as a predictor of colorectal neoplasia in ethnically diverse screening population. *Dig Dis Sci* 2010;55:2945–52.
18. Burke CA. Colonic complications of obesity. *Gastroenterol Clin North Am* 2010;39:47–55.
19. Tao S, Hoffmeister M, Brenner H. Development and validation of a scoring system to identify individuals at high risk for advanced colorectal neoplasms who should undergo colonoscopy screening. *Clin Gastroenterol Hepatol* 2014;12:478–85.
20. Gail MH, Benichou J. Validation studies on a model for breast cancer risk. *J Natl Cancer Inst* 1994;86:573–5.
21. Frikke-Schmidt R, Tybjaerg-Hansen A, Schnohr P, Jensen GB, Nordestgaard BG. Common clinical practice versus new PRIM score in predicting coronary heart disease risk. *Atherosclerosis* 2010;213:532–8.
22. Kamath PS, Kim WR. The model for end-stage liver disease (MELD). *Hepatology* 2007;45:797–805.
23. Merkel C, Zoli M, Siringo S, van Buuren H, Magalotti D, Angeli P, et al. Prognostic indicators of risk for first variceal bleeding in cirrhosis: a multicenter study in 711 patients to validate and improve the North Italian Endoscopic Club (NIEC) index. *Am J Gastroenterol* 2000;95:2915–20.
24. Beck DH, Smith GB, Pappachan JV, Millar B. External validation of the SAPS II, APACHE II and APACHE III prognostic models in South England: a multicentre study. *Intensive Care Med* 2003;29:249–56.
25. Provost F, Fawcett T. Analysis and visualization of classifier performance: comparison under imprecise class and cost distributions. In: *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining*. 1997. AAAI Press; 1997. p. 43–8.
26. Efron B, Tibshirani RJ (eds.). *An Introduction to Bootstrap*. New York: Chapman and Hall 1993.
27. 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.
28. Imperiale TF, Wagner DR, Lin CY, Larkin GN, Rogge JD, Ransohoff DF. Using risk for advanced proximal colonic neoplasia to tailor endoscopic screening for colorectal cancer. *Ann Intern Med* 2003;139:959–65.
29. Kaminski MF, Polkowski M, Kraszewska E, Rupinski M, Butruk E, Regula J. A score to estimate the likelihood of detecting advanced colorectal neoplasia at colonoscopy. *Gut* 2014;63:1112–9.
30. Yeoh KG, Ho KY, Chiu HM, Zhu F, Ching JY, Wu DC, Matsuda T, et al. The Asia-Pacific Colorectal Screening score: a validated tool that stratifies risk for colorectal advanced neoplasia in asymptomatic Asian subjects. *Gut* 2011;60:1236–41.
31. Cai QC, Yu ED, Xiao Y, Bai WY, Chen X, He LP, et al. Derivation and validation of a prediction rule for estimating advanced colorectal neoplasm risk in average-risk Chinese. *Am J Epidemiol* 2012;175:584–93.
32. Park HW, Hans S, Lee JS, Chang HS, Lee D, Choe JW. Risk stratification for advanced proximal colon neoplasm and individualized endoscopic screening for colorectal cancer by risk-scoring model. *Gut* 2014;63:1112–9.
33. Lipkin M, Blattner WA, Gardner EJ, Burt RW, Lynch H, Deschner E, et al. Classification and risk assessment of individuals with familial polyposis, Gardner's syndrome, and familial non-polyposis colon cancer from [3H]thymidine labeling patterns in colonic epithelial cells. *Cancer Res* 1984;44:4201–7.
34. Lieberman DA, Holub JL, Moravec MD, Eisen GM, Peters D, Morris CD, et al. Prevalence of colon polyps detected by colonoscopy screening in asymptomatic black and white patients. *JAMA* 2008;300:1417–22.
35. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol* 1974;66:688–701.
36. Hernan MA, Hernandez-Diaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol* 2002;155:176–84.
37. Giovannucci E, Ascherio A, Rimm EB, Colditz GA, Stampfer MJ, Willett WC. Physical activity, obesity, and risk for colon cancer and adenoma in men. *Ann Intern Med* 1995;122:327–34.
38. Giovannucci E, Colditz GA, Stampfer MJ, Willett WC. Physical activity, obesity, and risk of colorectal adenoma in women (United States). *Cancer Causes Control* 1996;7:253–63.
39. Schoenfeld P, Cash B, Flood A, Dobhan R, Eastone J, Coyle W, et al. Colonoscopic screening of average-risk women for colorectal neoplasia. *N Engl J Med* 2005;352:2061–8.
40. Kim DH, Lee SY, Choi KS, Lee HJ, Park SC, Kim J, et al. The usefulness of colonoscopy as a screening test for detecting colorectal polyps. *Hepato-gastroenterology* 2007;54:2240–2.
41. Rex DK, Lieberman D. ACG colorectal cancer prevention action plan: update on CT-colonography. *Am J Gastroenterol* 2006;101:1410–3.

# Cancer Epidemiology, Biomarkers & Prevention

## Development and Validation of a Clinical Score for Predicting Risk of Adenoma at Screening Colonoscopy

Aasma Shaukat, Timothy R. Church, Ryan Shanley, et al.

*Cancer Epidemiol Biomarkers Prev* 2015;24:913-920. Published OnlineFirst March 23, 2015.

**Updated version** Access the most recent version of this article at:  
doi:[10.1158/1055-9965.EPI-14-1321](https://doi.org/10.1158/1055-9965.EPI-14-1321)

**Cited articles** This article cites 38 articles, 5 of which you can access for free at:  
<http://cebp.aacrjournals.org/content/24/6/913.full#ref-list-1>

**Citing articles** This article has been cited by 2 HighWire-hosted articles. Access the articles at:  
<http://cebp.aacrjournals.org/content/24/6/913.full#related-urls>

**E-mail alerts** [Sign up to receive free email-alerts](#) related to this article or journal.

**Reprints and Subscriptions** To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at [pubs@aacr.org](mailto:pubs@aacr.org).

**Permissions** To request permission to re-use all or part of this article, use this link  
<http://cebp.aacrjournals.org/content/24/6/913>.  
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.