

Assessing individual risk for high-risk colorectal adenoma at first-time screening colonoscopy

Yin Cao^{1,2}, Bernard A. Rosner^{3,4}, Jing Ma^{2,3}, Rulla M. Tamimi^{2,3}, Andrew T. Chan^{3,5}, Charles S. Fuchs^{3,6}, Kana Wu^{1†} and Edward L. Giovannucci^{1,2,3†}

¹ Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA

² Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA

³ Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

⁴ Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA

⁵ Division of Gastroenterology, Massachusetts General Hospital, Boston, MA

⁶ Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA

Assessing risk of colorectal adenoma at first-time colonoscopy that are of higher likelihood of developing advanced neoplasia during surveillance could help tailor first-line colorectal cancer screening. We developed prediction models for high-risk colorectal adenoma (at least one adenoma ≥ 1 cm, or with advanced histology, or ≥ 3 adenomas) among 4,881 asymptomatic white men and 17,970 women who underwent colonoscopy as their first-time screening for colorectal cancer in two prospective US studies using logistic regressions. C-statistics and Hosmer–Lemeshow tests were used to evaluate discrimination and calibration. Ten-fold cross-validation was used for internal validation. A total of 330 (6.7%) men and 678 (3.8%) women were diagnosed with high-risk adenoma at first-time screening colonoscopy. The model for men included age, family history of colorectal cancer, BMI, smoking, sitting watching TV/VCR, regular aspirin/NSAID use, physical activity, and a joint term of multivitamin and alcohol. For women, the model included age, family history of colorectal cancer, BMI, smoking, alcohol, beef/pork/lamb as main dish, regular aspirin/NSAID, calcium, and oral contraceptive use. The C-statistic of the model for men was 0.67 and 0.60 for women (0.64 and 0.57 in cross-validation). Both models calibrated well. The predicted risk of high-risk adenoma for men in the top decile was 15.4% vs. 1.8% for men in the bottom decile (Odds Ratio [OR] = 9.41), and 6.6% vs. 2.1% for women (OR = 3.48). In summary, we developed and internally validated an absolute risk assessment tool for high-risk colorectal adenoma among the US population that may provide guidance for first-time colorectal cancer screening.

The decline in the incidence and mortality of colorectal cancer over the past two decades in the United States are largely attributable to colorectal cancer screening through prevention (identification and removal of polyps) and early detection.¹ Current screening guideline recommends adults with average risk of colorectal cancer should start screening at age 50 and at age 40 in general for people with family history.² Never-

theless, 35% of Americans aged 50–75 are not up-to-date with colorectal cancer screening.³ Despite a lack of consensus about a single best screening strategy, the demand for colonoscopy has surged in the past decade coincident with a decline in the use of sigmoidoscopy and fecal occult blood test. However, the majority of people who undergo colonoscopy do not harbor polyps.^{4,5} Finally, substantial uncertainty exists about whether national colonoscopy capacity would be sufficient for more widespread use of colonoscopy as the primary screening test.^{6–9} Therefore, tailoring colorectal cancer screening by identifying high risk individuals who should be encouraged to get a colonoscopy is critical for the efficient use of healthcare resources.^{4,10,11}

Risk assessment tools for colorectal cancer have been developed but the impact of previous endoscopy on subsequent cancer development was not fully assessed or was not designed to use in clinical settings because of requirement of periodic follow-up examinations.^{12–16} Alternatively, risk prediction for colorectal advanced adenoma may be more informative for screening and conservative for cancer prevention. According to the adenoma-carcinoma sequence,¹⁷ advanced adenomas (≥ 1 cm in diameter or with tubulovillous or villous histologic features or high-grade/severe

Key words: risk prediction, colorectal adenoma, colonoscopy, screening

[†]K.W. and E.L.G. contributed equally to this work.

Grant sponsor: NIH; **Grant number:** P01 CA87969, P01 CA55075, 1UM1 CA167552, R01 CA137178, K24 DK098311; **Grant sponsor:** Entertainment Industry Foundation through the National Colorectal Cancer Research Alliance. A.T.C. is a Damon Runyon Clinical Investigator.

DOI: 10.1002/ijc.29533

History: Received 14 Jan 2015; Accepted 12 Mar 2015; Online 26 Mar 2015

Correspondence to: Edward L. Giovannucci, Department of Nutrition and Epidemiology, Harvard T. H. Chan School of Public Health, 665 Huntington Avenue, Boston, MA 02115, USA, E-mail: egiovann@hsph.harvard.edu

What's new?

Population screening has dramatically reduced mortality rates for colorectal cancer (CRC) in the U.S. However, colonoscopy is expensive, and healthcare resources need to be more efficiently allocated. In this study, the authors developed and validated a comprehensive risk-assessment tool for high-risk colorectal adenoma. This tool, which includes a list of risk factors, may allow first-line colorectal cancer screening to identify people who are at increased risk. These people could then be encouraged to undergo colonoscopy, and to alter their behavior and lifestyle factors to lower their risk.

dysplasia), are the link between benign adenoma to colorectal cancer, and thus are considered as the primary target in screening endoscopies.^{4,18–21} In addition, multiplicity of adenomas (≥ 3) were associated with 2- to 5-fold increased risk of subsequent detection of advanced adenomas and cancer, and similar to advanced adenoma, surveillance at 3 years is recommended once identified.^{22,23}

A few risk prediction tools for advanced adenoma/neoplasia have been established recently among Asians and Europeans,^{24–28} but none were among US adults or restricted exclusively to first screening colonoscopy to aid risk stratification of first-time screening; we therefore sought to develop a risk assessment tool for high-risk colorectal adenoma (advanced adenoma or ≥ 3 adenomas) that can be implemented in clinical/general settings through evaluating a comprehensive list of risk factors, among white men and women who underwent colonoscopy as their first routine screening for colorectal cancer.

Material and Methods**Study population**

We utilized two large ongoing prospective cohort studies; the Health Professionals Follow-up Study (HPFS), a cohort study of 51,529 US male health professionals aged 40–75 at enrollment in 1986, and the Nurses' Health Study (NHS), a cohort study of 121,700 US female nurses aged 30–55 at enrollment in 1976. Participants have been mailed questionnaires at enrollment and 2 years thereafter to collect data on demographics, lifestyle factors, medical history, and disease outcomes and every 4 years to report update in dietary intake. The follow-up rates in both cohorts have been $>90\%$. The study protocol was approved by the institutional review boards of the Harvard School of Public Health and Brigham and Women's Hospital.

Between 1988 and 2002, both cohorts inquired whether the participants had a sigmoidoscopy or colonoscopy during the past 2 years and the indications for these procedures. Starting with the 2004 questionnaires, sigmoidoscopy and colonoscopy were assessed separately, and participants were also asked to provide information on when they had a sigmoidoscopy *versus* colonoscopy in any of the 2-year period in the past, including procedures before 1986.

We included asymptomatic men and women who were initially free of cancer (except non-melanoma skin cancer), ulcerative colitis, and who had reported colonoscopy as their first-time screening endoscopy for colorectal cancer between

1988–2008 (HPFS) and 1986–2008 (NHS). Because most of the participants were white (90% in HPFS and 97% in NHS), we restricted our analysis to white men and women. Individuals who reported occult or visible blood, diarrhea, constipation or abdominal pain as the indication(s) for the first colonoscopy were excluded.

Identification of cases and noncases

On each biennial questionnaire, participants were asked whether polyps had been diagnosed in the past 2 years. When a diagnosis was reported, we obtained informed consent to acquire medical records and pathology reports. Investigators blinded to any exposure information reviewed all records and extracted data on histological type, anatomic location and size of the polyps. If more than one adenoma was diagnosed, the subject was classified according to the adenoma of the largest size and most advanced histological characteristics.

Men or women diagnosed with at least one adenoma ≥ 1 cm in diameter, or with advanced histology (tubulovillous or villous histologic features or high grade or severe dysplasia), or ≥ 3 adenomas at their first colonoscopy were considered as high-risk adenomas and cases in this analysis. Non-cases included those diagnosed with other adenomas or hyperplastic polyps, or free of polyps at their first colonoscopy.

Assessment of risk factors

While selecting the set of risk factors for evaluation, we thoroughly reviewed existing epidemiological evidence as well as the simplicity of each risk factor assessment to ensure that our tool can be implemented in clinical and general settings. For risk factors with periodic updates in our cohorts, we extracted the most updated information before the report of the first colonoscopy without taking advantages of multiple measures over time, which may better reflect long-term exposure but is more impractical to implement.

We considered the following established/potential risk factors for advanced adenoma, multiple adenomas, and colorectal cancer including age, personal history of diabetes, family history of colorectal cancer in first-degree relatives, regular use of aspirin or NSAIDs (nonsteroidal anti-inflammatory drugs), multivitamin use, body mass index (BMI, kg m^{-2}), height (cm), calcium intake (mg day^{-1}) from supplement and food (total daily intake of milk, yogurt and calcium fortified orange juice multiplied by 300 mg),

supplemental vitamin D (IU day⁻¹), red meat intake in servings of beef/pork/lamb as a main dish, smoking in pack-years calculated as the product of packs of cigarettes smoked daily and the total years of smoking, alcohol intake (g day⁻¹) from total daily servings of beer, white wine, red wine and liquor, and sedentary behavior indicated by weekly hours of sitting watching TV/VCR. Physical activity level (low, moderate, high) was derived from weekly hours of brisk walking (≥ 3 miles/h) and time engaged in vigorous activities of ≥ 6 metabolic equivalents (running, jogging, bicycling, swimming, tennis, squash, racquetball and aerobics). We also evaluated oral contraceptive use, menopausal status and postmenopausal hormone use among women. The details of each risk factor assessment and validation were described previously.^{29–33} Because of the impracticality of using food frequency questionnaires to evaluate total dietary folate and vitamin D in clinical/general settings, we considered multivitamin use and supplemental vitamin D, which may capture part of the variations of the total intake of these two nutrients. Because the adverse effect of alcohol consumption on colorectal adenoma/cancer risk may be ameliorated by folic acid intake, we assessed the joint effect of alcohol and multivitamin use.^{34,35} When participants had missing information on a certain risk factor, we carried information assessed from the previous questionnaire. Participants with missing information after carrying forward were excluded.

Statistical analysis

For each risk factor, we first assessed its association with high-risk adenoma adjusting for 5-year age group by logistic regression. To assess dose-response, we compared the model fit between the models with continuous vs. categorical risk factor using log likelihood ratio test. We tested for statistical interaction by adding a cross-product term of the two variables to a model along with the main effects and compared to the model with only main effects using a log likelihood ratio test.

Stepwise multiple logistic regressions with $p_{\text{entry}} = 0.15$ and $p_{\text{stay}} = 0.20$ were used to build the prediction model in men and women, respectively. Odds ratios (ORs) and 95% confidence intervals (CI) of high-risk colorectal adenoma were presented. Predicted risk of high-risk colorectal adenoma was calculated by $\exp(\beta_0 + \sum \beta_i X_i) / (1 + \exp(\beta_0 + \sum \beta_i X_i))$, where β_0 was the intercept, and β_i was the regression coefficient for risk factor X_i .

Receiver-operating characteristic (ROC) curves were plotted for the final models as well as the models with age, and age and family history of colorectal cancer to evaluate the impact of adding modifiable risk factors to the current risk stratification criterion. C-statistics and age-adjusted C-statistics³⁶ which capture the predictive ability of the model without the impact of age were calculated to evaluate the discriminatory ability. Because of a limited sample size and overly pessimistic estimates of performance with large variability by split-sample method, as well as complex computation involved in bootstrapping, ten-fold cross-validation was

performed to validate the models internally.³⁷ We first randomly divided the data into 10 subsets, fitted the final model in 90% of the data (training set) and estimated the predictive risk of high-risk adenoma in the remaining 10% of the data (test set). We repeated this procedure for all the 10 subsets, and calculated the average C-statistics in the 10 test sets.

We assessed calibration by Hosmer–Lemeshow (HL) goodness-of-fit test comparing observed and predicted probabilities across deciles of predicted risk. p_{HL} values > 0.05 indicate adequate calibration.

As sensitivity analyses, we examined whether using continuous instead of categorical variables and a unisex model would improve discrimination. All the statistical analyses were conducted using SAS version 9.3 (SAS Institute, Cary, NC).

Results

A total of 4,881 men and 17,970 women had completed colonoscopy as their first endoscopy for routine colorectal cancer screening by 2008, and 330 (6.7%) men and 678 (3.8%) women were diagnosed with high-risk adenoma (Table 1).

In the final risk model for men, factors positively associated with risk of high-risk adenoma included age, family history of colorectal cancer, BMI, pack-years of smoking and time spent sitting watching TV/VCR (Table 2). Factors associated with reduced risk included regular use of aspirin/NSAIDs and moderate to high levels of physical activity. The risk model also included a joint term of alcohol intake and multivitamin use, with the highest risk observed among men who never took multivitamin and consumed ≥ 30 g of alcohol per day compared to men who never took multivitamin and consumed < 5 g of alcohol per day.

The risk model for women included the following factors that were positively associated with risk of high-risk adenoma: age, family history of colorectal cancer, BMI, smoking pack-years, alcohol intake, beef/pork/lamb as main dish; and factors that were inversely associated with risk of high-risk adenoma, including regular aspirin or NSAID use, calcium intake from supplement and food (milk, yogurt and orange juice), and oral contraceptive use (Table 2).

The C-statistic of the risk model for men was 0.67 (0.63 after adjusting for 5-year age group), and 0.60 (age-adjusted 0.59) for women (Table 3). C-statistics were not improved if continuous variables were used (0.65 for men and 0.59 for women) or when a unisex model with interactions was built combining data from men and women (0.64 for men and 0.59 for women).

In 10-fold cross-validation, the average C-statistic across the 10 test sets was 0.64 for men, and 0.57 for women. The discriminatory power of the model was higher among men younger than 65 (C-statistic = 0.69), and was less satisfactory among men aged 65 and above (C-statistic = 0.60); but was similar for women aged < 65 and 65 and above.

Significant improvement in C-statistics (overall, age-specific, age-adjusted overall and by type and location of

Table 1. Age-standardized characteristics of participants at first-time screening colonoscopy, HPFS (1988–2008) and NHS (1986–2008)

Characteristic	Men		Women	
	High-risk adenoma (n = 330)	Noncases (n = 4,551)	High-risk adenoma (n = 678)	Noncases (n = 17,292)
Age (yrs) ¹	65.6 (7.3)	63.1 (8.1)	65.0 (7.3)	64.2 (7.2)
Personal history of diabetes (%)	8.0	6.5	7.0	7.0
Family history of colorectal cancer (%)	20.6	15.7	28.6	22.0
Height (cm)	179 (6)	179 (7)	164 (6)	164 (6)
BMI (kg m ⁻²)	27.0 (3.6)	26.4 (3.7)	27.0 (5.3)	26.6 (5.2)
Postmenopausal hormone use (%)				
Never			25.9	22.9
Past			28.9	35.4
Current			40.6	37.0
Premenopausal (%)			4.6	4.8
Former or current oral contraceptive user (%)			52.6	56.2
Regular aspirin user (%)	35.7	39.7	36.8	39.9
Regular NSAID user (%)	17.7	24.4	35.3	41.5
Multivitamin				
Never (%)	25.0	18.4	2.9	3.0
Past (%)	17.9	20.6	37.8	29.8
Current (%)	57.1	61.0	59.2	67.2
Years of use	12.4 (7.8)	12.8 (8.1)	8.3 (6.6)	8.2 (6.2)
Calcium intake (mg day ⁻¹) ²	450 (437)	510 (465)	845 (598)	911 (608)
Vitamin D from supplement (IU day ⁻¹)	192 (235)	214(248)	238 (253)	288 (275)
Beef/pork/lamb as main dish, servings/wk	1.8 (1.4)	1.7 (1.5)	1.5 (1.3)	1.4 (1.3)
Former or current smoker (%)	61.8	54.0	58.4	54.0
Pack-years of smoking	23.4 (19.8)	17.9 (17.2)	23.1 (18.4)	20.5 (17.7)
Alcohol (g day ⁻¹)	12.7 (14.9)	12.1 (15.4)	6.6 (11.9)	5.8 (10.0)
Brisk walking (h wk ⁻¹) ³	1.6 (4.2)	1.9 (4.0)	0.8 (1.9)	0.9 (2.2)
Vigorous activity (h wk ⁻¹) ³	1.4 (2.3)	2.0 (3.6)	1.1 (4.0)	0.9 (2.3)
Physical activity ³				
Low (%)	43.9	33.8	48.2	49.5
Moderate (%)	28.7	32.6	37.7	34.9
High (%)	27.3	33.6	14.1	15.5
Sitting watching TV/VCR (h day ⁻¹)	1.5 (1.2)	1.4 (1.2)	1.8 (1.6)	1.8 (1.6)

¹All values were calculated from the most recent questionnaire before the first colonoscopy, and other than age, have been directly standardized to age distribution of all the participants. Mean (SD) was presented for continuous variables.

²Calcium intake was the sum of calcium intake from supplement and food. Calcium from food was derived by multiplying the total daily intake of milk, yogurt and fortified orange juice by 300 mg.

³Physical activity was defined by levels of brisk walking (walking speed ≥ 3 miles hr⁻¹) and vigorous activity (activities with ≥ 6 metabolic equivalents including running, jogging, bicycling, swimming, tennis, squash, racquetball, and aerobics). Low: 0–4.9 hrs of brisk walking per week and no vigorous activity; moderate: 0–4.9 hrs of brisk walking and 0.1–2.9 hrs of vigorous activity; high: ≥ 5 hrs of brisk walking or ≥ 3 hrs of vigorous activity.

high-risk adenoma) were observed in both men and women when our models were compared to the models with only age and family history (Table 3), suggesting the importance of adding modifiable lifestyle risk factors to the risk stratification of high-risk adenoma. ROC curves in Figure 1 indi-

cate that our models were also superior to models with age alone.

In both men and women, our models were better at predicting ≥ 3 adenomas (C-statistic = 0.68 for men and 0.65 for women), compared to large adenomas and adenomas of

Table 2. Risk prediction models for high-risk colorectal adenoma at first-time screening colonoscopy

Characteristic	Men				Women			
	Case/non-case (330/4551)	OR	(95% CI)	<i>p</i> *	Case/noncase (678/17,292)	OR	(95% CI)	<i>p</i> *
Age (yrs)								
<55	17/639	1.00	–	<0.001	61/1,550	1.00	–	0.10
55–59	62/1,195	2.00	(1.15,3.47)		105/3,347	0.85	(0.61,1.17)	
60–64	84/1,019	3.08	(1.80,5.29)		172/4,781	0.96	(0.71,1.31)	
65–69	73/750	3.42	(1.97,5.91)		165/3,882	1.13	(0.83,1.54)	
≥70	94/948	3.64	(2.13,6.23)		175/3,732	1.26	(0.92,1.73)	
Family history of colorectal cancer	65/719	1.43	(1.07,1.91)	0.01	193/3,814	1.41	(1.18,1.67)	<0.001
BMI (kg m⁻²)¹								
Quartile 1	54/1,192	1.00	–	<0.001	149/4,349	1.00	–	0.14
Quartile 2	90/1,172	1.59	(1.12,2.26)		167/4,338	1.14	(0.91,1.43)	
Quartile 3	87/1,050	1.69	(1.18,2.41)		181/4,330	1.23	(0.99,1.54)	
Quartile 4	99/1,137	1.73	(1.22,2.47)		181/4,275	1.29	(1.03,1.61)	
Aspirin (regular vs. non-regular)	121/1,801	0.78	(0.62,0.99)	0.03	253/6,901	0.88	(0.75,1.04)	0.15
NSAIDs (regular vs. non-regular)	64/1,110	0.70	(0.53,0.93)	0.02	238/7,186	0.77	(0.65,0.91)	0.001
Smoking, pack-years								
Never smoked	121/2,101	1.00	–	0.008	283/7,950	1.00	–	<0.001
1–4	36/646	0.91	(0.62,1.35)		61/1,910	0.93	(0.70,1.24)	
5–19	66/839	1.29	(0.94,1.77)		132/3,380	1.11	(0.90,1.37)	
20–39	69/674	1.46	(1.06,2.01)		122/2,683	1.25	(1.00,1.55)	
≥40	38/291	1.64	(1.10,2.45)		80/1,369	1.50	(1.15,1.95)	
Physical activity²								
Low	145/1,534	1.00	–	0.06				
Moderate	98/1,486	0.81	(0.62,1.07)					
High	87/1,531	0.70	(0.53,0.94)					
Sitting watching TV/VCR (h day⁻¹)								
<0.5	78/1,267	1.00	–	0.004				
0.5–<2	110/1,831	1.03	(0.76,1.39)					
≥2	142/1,453	1.49	(1.11,1.99)					
Alcohol*Multivitamin								
<5 g day ⁻¹ , never used multivitamin	25/392	1.00	–	0.08				
<5 g day ⁻¹ , ever used multivitamin	104/1,602	1.03	(0.65,1.62)					
5–29 g day ⁻¹ , never used multivitamin	34/356	1.44	(0.84,2.49)					
5–29 g day ⁻¹ , ever used multivitamin	121/1,675	1.19	(0.75,1.88)					
≥30 g day ⁻¹ , never used multivitamin	15/92	2.32	(1.15,4.66)					
≥30 g day ⁻¹ , ever used multivitamin	31/434	1.02	(0.59,1.79)					
Calcium (mg day⁻¹)³								

Table 2. Risk prediction models for high-risk colorectal adenoma at first-time screening colonoscopy (Continued)

Characteristic	Men				Women			
	Case/non-case (330/4551)	OR	(95% CI)	<i>p</i> *	Case/noncase (678/17,292)	OR	(95% CI)	<i>p</i> *
<300					139/2,881	1.00	–	0.07
300–599					132/3,173	0.88	(0.69,1.13)	
≥600					407/11,238	0.80	(0.66,0.98)	
Oral contraceptive (ever vs. never)					346/9,728	0.89	(0.76,1.05)	0.03
Alcohol (≥30 vs.<30 g day ⁻¹)					40/712	1.32	(0.95,1.85)	0.09
Beef/pork/lamb as main dish, servings								
<2/mo					163/4,506	1.00	–	0.05
2/mo–<2/wk					286/7,558	1.08	(0.88,1.31)	
≥2/wk					229/5,228	1.26	(1.02,1.55)	

**p* values for stepwise selection.

¹BMI quartile cutoff points in men: 24.1, 25.9, 28.2 kg m⁻²; in women: 22.9, 25.8, 29.3 kg m⁻².

²Physical activity was defined by levels of brisk walking (walking speed ≥3 miles/hr) and vigorous activity (activities with ≥6 metabolic equivalents including running, jogging, bicycling, swimming, tennis, squash, racquetball and aerobics). Low: 0–4.9 hrs of brisk walking per week and no vigorous activity; moderate: 0–4.9 hrs of brisk walking and 0.1–2.9 hrs of vigorous activity; high: ≥5 hrs of brisk walking or ≥3 hrs of vigorous activity.

³Calcium intake was the sum of calcium intake from supplement and food. Calcium from food was derived by multiplying the total daily intake of milk, yogurt and fortified orange juice by 300 mg.

Table 3. C-statistics of the risk prediction models for high-risk colorectal adenoma compared with models with age and family history

	Men				Women			
	Case/ noncase	Age + family history	Final model	<i>p</i>	Case/ noncase	Age + family history	Final model	<i>p</i>
		C	C			C	C	
Overall	330/4,551	0.61	0.67	<0.001	678/17,292	0.56	0.60	<0.001
By age								
<65	163/2,853	0.61	0.69	<0.001	338/9,678	0.54	0.58	0.01
≥65	167/1,698	0.52	0.60	0.01	340/7,614	0.55	0.60	0.003
Age-adjusted¹	330/4,551	0.53	0.63	<0.001	678/17,292	0.53	0.59	<0.001
By type¹								
≥1 cm in diameter	242/4,551	0.52	0.61	<0.001	508/17,292	0.53	0.59	<0.001
Advanced histology ²	150/4,551	0.56	0.63	0.01	364/17,292	0.54	0.60	<0.001
≥3 adenomas	126/4,551	0.53	0.68	<0.001	140/17,292	0.55	0.65	<0.001
By location¹								
Rectal	68/4,551	0.47	0.62	0.001	130/17,292	0.53	0.59	0.03
Distal	178/4,551	0.54	0.64	<0.001	378/17,292	0.53	0.61	<0.001
Proximal	124/4,551	0.54	0.65	<0.001	226/17,292	0.55	0.59	0.07

¹Adjusted for 5-year age groups

²With villous components (tubulovillous or villous), or with high-grade or severe dysplasia.

advanced histology (Table 3). The C-statistics were similar across sub-sites of the colorectum (*i.e.*, proximal, distal and rectal high-risk adenomas).

Both models calibrated well across deciles of predicted risk ($p_{HL} = 0.48$ for men and 0.96 for women) (Table 4). The

predicted risk of high-risk adenoma for men in the top decile was 15.4% compared to 1.8% for men in the bottom decile (OR = 9.41; 95% CI 4.46–19.8). For women, the predicted risk in the top decile was 6.6% compared to 2.1% for women in the bottom decile (OR = 3.48; 95% CI 2.38–5.08). The

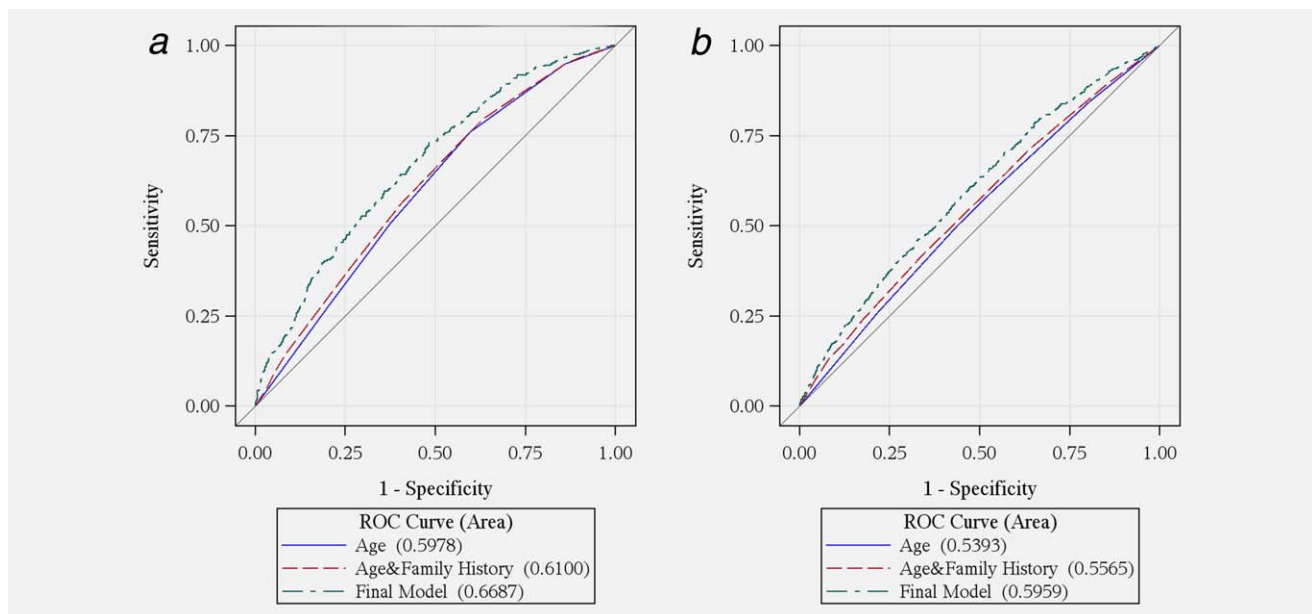


Figure 1. ROC curves comparing models with age, age and family history, and the final prediction models of high-risk adenoma at first-time screening colonoscopy. (a) Men; (b) Women. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

models also calibrated well among people under age 65 and aged 65 and above.

Discussion

Using data of two large prospective cohort studies, we extensively evaluated established and potential risk factors for colorectal neoplasia to develop separate risk prediction models for high-risk adenoma among 22,851 asymptomatic white men and women who underwent colonoscopy as the first routine screening for colorectal cancer. The models performed adequately in terms of discriminatory accuracy and goodness-of-fit. To the best of our knowledge, this set of prediction models is the first comprehensive tool for risk assessment of high-risk adenoma among asymptomatic US adults.

The unique strengths of our study are (i) the use of high-risk adenoma diagnosed at first-time screening colonoscopy as the primary study endpoint, which is not influenced by previous screening history, provide direct information to those who never had a colorectal cancer screening, and is more conservative and informative for cancer prevention as advanced adenomas are precursors of the majority of colorectal cancers and the target of screening colonoscopies, and multiple adenomas infer higher risk of detection of advanced adenoma at surveillance colonoscopy; (ii) the estimation of absolute risk, a more useful parameter in clinical decision than relative risks alone; (iii) the attempt to develop a simple risk calculator that could be conveniently used in general settings by taking into account the feasibility of each risk factor assessment; (iv) the ability to consider a wide variety of prospectively collected data on established and suspected risk factors of colorectal neoplasia for model development; (v) the comparison to the models with only age and family history

to evaluate the impact of adding modifiable risk factors to the current screening risk stratification criterion.

Previous models established in other western countries^{24,27,28} did not address the above issues at the same time. Most importantly, these models were developed in mixed population with first and surveillance colonoscopies. Risk of recurrent adenomas largely depends on pathologic findings in the first screening colonoscopy and could not help address whether the baseline screening colonoscopy should be initiated. As risk factors as well as strengths of associations may also vary for prevalent and recurrent adenomas, these models need to be improved. Discrimination of the model by Tao *et al.* will be less optimal in discrimination (C-statistic 0.67) when restricted to first colonoscopies because previous colonoscopy and polyps history (relative risk [RR] = 0.4 and 3.2, respectively) were the strongest risk factors included in the model.²⁴ Moreover, established and prevalent risk factors, such as aspirin/NSAID use was not included in Kaminski *et al.*²⁸ and Betes *et al.*²⁷

Gender-specific absolute risk prediction models for colorectal cancer have been developed utilizing two case-control studies and SEER data, and validated in an external population with modest C-statistics (0.61 for both men and women).^{13,14} However, risk factors were assessed retrospectively, history of endoscopies were only accessed crudely up to previous 10 years with missing data, and a list of establish/potential risk factors were not evaluated in the model development (*e.g.*, history of diabetes,³⁸ calcium intake³⁹ and height⁴⁰).

The C-statistics for our models (0.67 for men and 0.60 for women) were modest but comparable to other cancer risk models, *e.g.*, 0.58–0.63 for breast cancer,^{41–44} 0.69 for lung cancer,⁴⁵ 0.59–0.60 for ovarian cancer.^{41,46}

Table 4. Calibration of risk prediction models for high-risk colorectal adenoma overall and by age across deciles of predicted risk

	Decile of predicted risk of high-risk colorectal adenoma										p_{HL}^1	OR (95% CI) ²
	1	2	3	4	5	6	7	8	9	10		
Men												
Overall (Case/Non-case = 330/4,551)												
Observed	8	12	20	27	24	38	34	38	63	66		
Expected	8.7	14.1	18.5	22.5	26.6	31.4	36.9	43.4	53.0	75.0		
Observed risk	1.6%	2.5%	4.1%	5.5%	4.9%	7.8%	7.0%	7.8%	12.9%	13.6%		
Predicted Risk	1.8%	2.9%	3.8%	4.6%	5.5%	6.4%	7.6%	8.9%	10.8%	15.4%	0.48	9.41 (4.46,19.8)
Age<65 (case/noncase = 163/2,853)												
Observed	4	6	10	9	15	17	14	18	22	48		
Expected	4.6	7.0	9.1	11.1	12.8	15.3	17.7	21.2	25.7	38.6		
Observed risk	1.3%	2.0%	3.3%	3.0%	5.0%	5.6%	4.7%	5.9%	7.3%	15.9%		
Predicted Risk	1.5%	2.3%	3.0%	3.7%	4.3%	5.0%	5.9%	7.0%	8.6%	12.8%	0.14	14.0 (4.99,39.4)
Age≥65 (case/noncase = 167/1,698)												
Observed	6	12	16	18	14	16	16	20	15	34		
Expected	6.8	9.2	11.0	12.8	14.6	16.3	18.5	20.8	23.9	33.2		
Observed risk	3.2%	6.4%	8.6%	9.7%	7.5%	8.6%	8.5%	10.8%	8.1%	18.2%		
Predicted Risk	3.6%	4.9%	5.9%	6.9%	7.8%	8.7%	9.8%	11.2%	12.8%	17.7%	0.28	6.67 (2.73,16.3)
Women												
Overall (case/non-case = 678/17,292)												
Observed	36	44	48	64	62	72	67	78	88	119		
Expected	37.6	45.6	51.2	56.1	61.0	66.1	72.0	79.7	90.7	118		
Observed risk	2.0%	2.5%	2.7%	3.6%	3.5%	4.0%	3.7%	4.3%	4.9%	6.6%		
Predicted Risk	2.1%	2.5%	2.8%	3.1%	3.4%	3.7%	4.0%	4.4%	5.0%	6.6%	0.96	3.48 (2.38,5.08)
Age<65 (case/non-case = 338/9,678)												
Observed	26	17	28	26	31	40	35	41	36	58		
Expected	19.7	23.2	25.9	28.2	30.8	32.9	35.9	39.5	44.7	57.2		
Observed risk	2.6%	1.7%	2.8%	2.6%	3.1%	4.0%	3.5%	4.1%	3.6%	5.8%		
Predicted risk	2.0%	2.3%	2.6%	2.8%	3.1%	3.3%	3.6%	3.9%	4.5%	5.7%	0.48	2.32 (1.45,3.71)
Age≥65 (case/non-case = 340/7,614)												
Observed	17	24	23	33	29	31	44	40	40	59		
Expected	20.0	23.8	26.3	28.5	30.7	33.0	36.0	39.5	44.6	57.6		
Observed risk	2.1%	3.0%	2.9%	4.2%	3.6%	3.9%	5.5%	5.0%	5.0%	7.4%		
Predicted Risk	2.5%	3.0%	3.3%	3.6%	3.8%	4.2%	4.5%	5.0%	5.6%	7.2%	0.85	3.66 (2.12,6.34)

¹ p_{HL} : p value for Hosmer–Lemeshow goodness-of-fit test; $p > 0.05$ indicates adequate fit.

²OR of high-risk adenoma for the top decile compared with the bottom decile of predicted prevalence.

Almost all of the factors in our final risk models have been consistently reported as risk factors for colorectal adenoma and cancer.^{34,35,39,47–55} Sedentary behaviors as indicated by TV viewing are less studied⁵⁶; however, independent of physical activity, prolonged TV watching has been associated

with increased risk of obesity,⁵⁷ a known risk factor for colorectal carcinogenesis.⁴⁹ We built separate models for men and women because risk factors or the strengths of association for high-risk adenoma/colorectal cancer may vary by gender.^{49,50,54} Consistent with previous findings,^{10,49} we

found that influence of BMI and age were stronger in men than in women. For factors that emerged in the prediction model one sex but not the other may reflect a true etiologic association that is stronger in one group than the other (e.g., physical activity is a slighter stronger risk factor for men⁵⁴); or our simplified definition of risk factors in one group may not fully capture long-term exposure variations as compared to the other group (e.g., calcium intake from supplement, milk, yogurt and fortified orange juice may capture less long-term variations of total calcium intake among men). Our models should be interpreted from the standpoint of clinical risk prediction not etiologic research.

Our study has limitations. The absolute risk of high-risk adenoma observed (6.7% in men and 3.8% in women) was comparable to other studies^{4,10}; however, we were limited in power to fully assess the impact of extreme risk factors because health professionals are slightly more health-conscious than the general US population (e.g., prevalence of current smoking was 4% in the HPFS compared to national average of 22%⁵⁸). Additionally, our prediction was most applicable to white men and women, calibration in additional studies especially among other races were necessary before use in general settings. Finally, colonoscopy or sigmoidoscopy

was self-reported. However, misclassification was unlikely as colonoscopy requires sedation.

In summary, we developed and evaluated a set of risk stratification tool that can be used to estimate current risk of high-risk colorectal adenoma and provide guidance regarding first-time colorectal cancer screening. Our models may encourage high-risk people to undertake colonoscopy, and will be particularly valuable in resource limited settings. Low-risk individuals should discuss screening alternatives (e.g., fecal occult blood test) with their physicians. The models could also help people target behaviors to lower their risk of colorectal neoplasia. A web/mobile based risk calculator/nomogram may increase the feasibility.

Acknowledgements

The authors thank participants and staff of the Health Professionals Follow-up Study and the Nurses' Health Study for their valuable contributions and the state cancer registries for their help: Alabama, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, Florida, Georgia, Idaho, Illinois, Indiana, Iowa, Kentucky, Louisiana, Maine, Maryland, Maine, Michigan, Nebraska, New Hampshire, New Jersey, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, Tennessee, Texas, Virginia, Washington, Wyoming. They thank Mr. Scott G. Smith, from Department of Nutrition, Harvard School of Public Health, for his contribution in data management.

References

- Edwards BK, Ward E, Kohler BA, et al. Annual report to the nation on the status of cancer, 1975–2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. *Cancer* 2010;116:544–73.
- Levin B, Lieberman DA, McFarland B, 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.
- Klabunde CN, Joseph DA, King JB, et al. Vital signs: colorectal cancer screening test use—United States, 2012. *MMWR Morbidity and Mortality Weekly Rep* 2013;62:881–8.
- Regula J, Rupinski M, Kraszewska E, et al. Colonoscopy in colorectal-cancer screening for detection of advanced neoplasia. *N Engl J Med* 2006;355:1863–72.
- Schroy PC, Coe A, Chen CA, et al. Prevalence of advanced colorectal neoplasia in white and black patients undergoing screening colonoscopy in a safety-net hospital. *Ann Intern Med* 2013;159:13.
- Vijan S, Inadomi J, Hayward RA, et al. Projections of demand and capacity for colonoscopy related to increasing rates of colorectal cancer screening in the United States. *Aliment Pharmacol Therap* 2004;20:507–15.
- Seeff LC, Manninen DL, Dong FB, et al. Is there endoscopic capacity to provide colorectal cancer screening to the unscreened population in the United States? *Gastroenterology* 2004;127:1661–9.
- Steinwachs D, Allen JD, Barlow WE, et al. National institutes of health state-of-the-science conference statement: enhancing use and quality of colorectal cancer screening. *Ann Intern Med* 2010;152:663–7.
- Ballew C, Lloyd BG, Miller SH. Capacity for colorectal cancer screening by colonoscopy, montana, 2008. *Am J Prevent Med* 2009;36:329–32.
- Ferlitsch M, Reinhart K, Pramhas S, et al. Sex-specific prevalence of adenomas, advanced adenomas, and colorectal cancer in individuals undergoing screening colonoscopy. *JAMA J Am Med Assoc* 2011;306:1352–8.
- Brenner H, Hoffmeister M, Stegmaier C, et al. Risk of progression of advanced adenomas to colorectal cancer by age and sex: estimates based on 840,149 screening colonoscopies. *Gut* 2007;56:1585–9.
- Colditz GA, Atwood KA, Emmons K, et al. Harvard report on cancer prevention volume 4: Harvard cancer risk index. Risk index working group, Harvard center for cancer prevention. *Cancer Causes Control CCC* 2000;11:477–88.
- Freedman AN, Slattery ML, Ballard-Barbash R, et al. Colorectal cancer risk prediction tool for white men and women without known susceptibility. *J Clin Oncol Off J Am Soc Clin Oncol* 2009;27:686–93.
- 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;27:694–8.
- 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;170:863–72.
- Driver JA, Gaziano JM, Gelber RP, et al. Development of a risk score for colorectal cancer in men. *Am J Med* 2007;120:257–63.
- Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell* 1990;61:759–67.
- Atkin WS, Morson BC, Cuzick J. Long-term risk of colorectal cancer after excision of rectosigmoid adenomas. *N Engl J Med* 1992;326:658–62.
- Winawer SJ, Zauber AG. The advanced adenoma as the primary target of screening. *Gastrointest Endosc Clin N Am* 2002;12:1–9.
- Imperiale TF, Wagner DR, Lin CY, et al. Using risk for advanced proximal colonic neoplasia to tailor endoscopic screening for colorectal cancer. *Ann Intern Med* 2003;139:959–65.
- Winawer S, Fletcher R, Rex D, et al. Colorectal cancer screening and surveillance: clinical guidelines and rationale—update based on new evidence. *Gastroenterology* 2003;124:544–60.
- Lieberman DA, Weiss DG, Harford WV, et al. Five-year colon surveillance after screening colonoscopy. *Gastroenterology* 2007;133:1077–85.
- Lieberman DA, Rex DK, Winawer SJ, et al. Guidelines for colonoscopy surveillance after screening and polypectomy: a consensus update by the US multi-society task force on colorectal cancer. *Gastroenterology* 2012;143:844–57.
- 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 Off Clin Practice J Am Gastroenterol Assoc* 2014;12:478–85.
- Yeoh KG, Ho KY, Chiu HM, 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.
- Cai QC, Yu ED, Xiao Y, 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.

27. Betes M, Munoz-Navas MA, Duque JM, et al. Use of colonoscopy as a primary screening test for colorectal cancer in average risk people. *Am J Gastroenterol* 2003;98:2648–54.
28. Kaminski MF, Polkowski M, Kraszewska E, et al. A score to estimate the likelihood of detecting advanced colorectal neoplasia at colonoscopy. *Gut* 2014;63:1112–9.
29. Pan A, Sun Q, Bernstein AM, et al. Red meat consumption and risk of type 2 diabetes: 3 cohorts of US adults and an updated meta-analysis. *Am J Clin Nutr* 2011;94:1088–96.
30. Feskanich D, Rimm EB, Giovannucci EL, et al. Reproducibility and validity of food intake measurements from a semiquantitative food frequency questionnaire. *J Am Dietetic Assoc* 1993;93:790–6.
31. Rimm EB, Giovannucci EL, Stampfer MJ, et al. Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. *Am J Epidemiol* 1992;135:1114–26; discussion 27–36.
32. Chasan-Taber S, Rimm EB, Stampfer MJ, et al. Reproducibility and validity of a self-administered physical activity questionnaire for male health professionals. *Epidemiology* 1996;7:81–6.
33. Wolf AM, Hunter DJ, Colditz GA, et al. Reproducibility and validity of a self-administered physical activity questionnaire. *Int J Epidemiol* 1994;23:991–9.
34. Giovannucci E, Rimm EB, Ascherio A, et al. Alcohol, low-methionine–low-folate diets, and risk of colon cancer in men. *J Natl Cancer Inst* 1995;87:265–73.
35. Giovannucci E, Stampfer MJ, Colditz GA, et al. Folate, methionine, and alcohol intake and risk of colorectal adenoma. *J Natl Cancer Inst* 1993;85:875–84.
36. 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;65:188–97.
37. Steyerberg EW, Harrell FE, Borsboom GJJM, et al. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 2001;54:774–81.
38. Larsson SC, Orsini N, Wolk A. Diabetes mellitus and risk of colorectal cancer: a meta-analysis. *J Natl Cancer Inst* 2005;97:1679–87.
39. Cho E, Smith-Warner SA, Spiegelman D, et al. Dairy foods, calcium, and colorectal cancer: a pooled analysis of 10 cohort studies. *J Natl Cancer Inst* 2004;96:1015–22.
40. Engeland A, Tretli S, Austad G, et al. Height and body mass index in relation to colorectal and gallbladder cancer in two million Norwegian men and women. *Cancer Cause Control* 2005;16:987–96.
41. 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:e1001492.
42. 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* 2001;93:358–66.
43. 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;152:950–64.
44. Barlow WE, White E, Ballard-Barbash R, et al. Prospective breast cancer risk prediction model for women undergoing screening mammography. *J Natl Cancer Inst* 2006;98:1204–14.
45. Cronin KA, Gail MH, Zou Z, et al. Validation of a model of lung cancer risk prediction among smokers. *J Natl Cancer Inst* 2006;98:637–40.
46. Rosner BA, Colditz GA, Webb PM, et al. Mathematical models of ovarian cancer incidence. *Epidemiology* 2005;16:508–15.
47. Cho E, Smith-Warner SA, Ritz J, et al. Alcohol intake and colorectal cancer: a pooled analysis of 8 cohort studies. *Ann Inter Med* 2004;140:603–13.
48. Giovannucci E, Ascherio A, Rimm EB, et al. Physical activity, obesity, and risk for colon cancer and adenoma in men. *Ann Inter Med* 1995; 122:327–34.
49. Moghaddam AA, Woodward M, Huxley R. Obesity and risk of colorectal cancer: a meta-analysis of 31 studies with 70,000 events. *Cancer Epidemiol Biomarkers Prevent Publ Am Assoc Cancer Res Am Soc Prevent Oncol* 2007; 16:2533–47.
50. Bosetti C, Bravi F, Negri E, et al. Oral contraceptives and colorectal cancer risk: a systematic review and meta-analysis. *Hum Reprod Update* 2009;15:489–98.
51. Botteri E, Iodice S, Raimondi S, et al. Cigarette smoking and adenomatous polyps: a meta-analysis. *Gastroenterology* 2008;134:388–95.
52. Chan AT, Giovannucci EL, Schernhammer ES, et al. A prospective study of aspirin use and the risk for colorectal adenoma. *Ann Inter Med* 2004; 140:157–66.
53. Chan AT, Giovannucci EL, Meyerhardt JA, et al. Long-term use of aspirin and nonsteroidal anti-inflammatory drugs and risk of colorectal cancer. *JAMA J Am Med Assoc* 2005;294:914–23.
54. Wolin KY, Yan Y, Colditz GA. Physical activity and risk of colon adenoma: a meta-analysis. *Br J Cancer* 2011;104:882–5.
55. Wolin KY, Yan Y, Colditz GA, et al. Physical activity and colon cancer prevention: a meta-analysis. *Br J Cancer* 2009;100:611–6.
56. Schmid D, Leitzmann MF. Television viewing and time spent sedentary in relation to cancer risk: a meta-analysis. *J Natl Cancer Inst* 2014; 106. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24935969>.
57. Hu FB, Li TY, Colditz GA, et al. Television watching and other sedentary behaviors in relation to risk of obesity and type 2 diabetes mellitus in women. *JAMA J Am Med Assoc* 2003;289: 1785–91.
58. King B, Dube S, Kaufmann R, et al. Vital signs: current cigarette smoking among adults aged ≥ 18 years—United States, 2005–2010. *MMWR Morbidity Mortality Weekly Rep* 2011;60:1207–12.

THE DIFFERENCE OF BREAKTHROUGH MOMENTS

WITH AN INTEGRATED SOLUTION FOR
GROUNDBREAKING SINGLE-CELL RESEARCH

BD Accuri™ C6 Plus Personal Flow Cytometer

BD FACSCelesta™ Cell Analyzer

BD LSRFortessa™ X-20 Cell Analyzer

BD FACSMelody™ Cell Sorter

FlowJo™ Software

One of the largest portfolios of reagents

Discover more >

