



OPEN ACCESS

ORIGINAL ARTICLE

A score to estimate the likelihood of detecting advanced colorectal neoplasia at colonoscopy

Michał F Kaminski,^{1,2} Marcin Polkowski,^{1,2} Ewa Kraszewska,² Maciej Rupinski,^{1,2} Eugeniusz Butruk,³ Jarosław Reguła^{1,2}

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/gutjnl-2013-304965>).

¹Department of Gastroenterological Oncology, The Maria Skłodowska-Curie Memorial Cancer Centre and Institute of Oncology, Warsaw, Poland

²Department of Gastroenterology and Hepatology, Medical Centre for Postgraduate Education, Warsaw, Poland

³The Polish Foundation of Gastroenterology, Warsaw, Poland

Correspondence to

Professor Jarosław Reguła, Department of Gastroenterological Oncology, Institute of Oncology, Roentgen Street 5, Warsaw 02-781, Poland; jregula@coi.waw.pl

Received 25 March 2013

Revised 4 December 2013

Accepted 9 December 2013

Published Online First

2 January 2014

ABSTRACT

Objective This study aimed to develop and validate a model to estimate the likelihood of detecting advanced colorectal neoplasia in Caucasian patients.

Design We performed a cross-sectional analysis of database records for 40-year-old to 66-year-old patients who entered a national primary colonoscopy-based screening programme for colorectal cancer in 73 centres in Poland in the year 2007. We used multivariate logistic regression to investigate the associations between clinical variables and the presence of advanced neoplasia in a randomly selected test set, and confirmed the associations in a validation set. We used model coefficients to develop a risk score for detection of advanced colorectal neoplasia.

Results Advanced colorectal neoplasia was detected in 2544 of the 35 918 included participants (7.1%). In the test set, a logistic-regression model showed that independent risk factors for advanced colorectal neoplasia were: age, sex, family history of colorectal cancer, cigarette smoking ($p < 0.001$ for these four factors), and Body Mass Index ($p = 0.033$). In the validation set, the model was well calibrated (ratio of expected to observed risk of advanced neoplasia: 1.00 (95% CI 0.95 to 1.06)) and had moderate discriminative power (c-statistic 0.62). We developed a score that estimated the likelihood of detecting advanced neoplasia in the validation set, from 1.32% for patients scoring 0, to 19.12% for patients scoring 7–8.

Conclusions Developed and internally validated score consisting of simple clinical factors successfully estimates the likelihood of detecting advanced colorectal neoplasia in asymptomatic Caucasian patients. Once externally validated, it may be useful for counselling or designing primary prevention studies.

INTRODUCTION

The strength of evidence regarding the efficacy of colorectal cancer screening in reducing the incidence of colorectal cancer and associated mortality is increasing.^{1–2} Colorectal cancer screening is currently recommended in the European Union³; however, adherence to this recommendation is not sufficient.^{4–6} One of the most important barriers to screening is a lack of perceived risk of colorectal cancer among average-risk patients and primary care providers.^{7–8} The risk of colorectal cancer or advanced colorectal neoplasia varies with regard to several factors, including age,^{9–11} sex,^{10–12} family history of colorectal cancer,^{10–13} smoking,^{14–15} obesity,^{11–16} diabetes mellitus,¹⁷ long-term non-steroid anti-inflammatory drug use,^{15–18} diet^{15–19}

Significance of this study

What is already known about this subject?

- Colorectal cancer screening is currently recommended for average-risk population older than 50 years of age, but the adherence to this recommendation is generally not sufficient.
- One of the most important barriers to screening is low perceived risk of colorectal cancer among average-risk patients and primary care providers.
- Advanced colorectal neoplasia is not evenly distributed throughout the 'average-risk' screening population.

What are the new findings?

- The new score uses age, sex, family history of colorectal cancer, cigarette smoking and Body Mass Index to estimate the likelihood of detecting advanced colorectal neoplasia in asymptomatic Caucasian patients.
- The score estimated the likelihood of detecting advanced colorectal neoplasia in so-called 'average-risk' asymptomatic Caucasian patients from 1.3% to 19.1%.

How might it impact on clinical practice in the foreseeable future?

- The score may increase colorectal cancer risk awareness and help healthcare providers to encourage people to get screened by adhering to existing national screening programmes.
- By identifying patients at high likelihood of detecting advanced colorectal neoplasia, the score may help to target primary prevention interventions.



Open Access
Scan to access more
free content



CrossMark

To cite: Kaminski MF, Polkowski M, Kraszewska E, et al. *Gut* 2014;**63**: 1112–1119.

scoring system that estimates the likelihood of detecting advanced colorectal neoplasia in asymptomatic patients.

METHODS

Study design and oversight

We performed a cross-sectional analysis of database records for 40-year-old to 66-year-old patients who entered the national colonoscopy screening programme for colorectal cancer in Poland, from January 2007 through December 2007. The database contained demographic data, colonoscopy and histopathology results, follow-up information, and the results of an epidemiological questionnaire on potential risk factors for advanced colorectal neoplasia from 73 screening centres throughout Poland.

The research proposal was reviewed by the Research Ethical Committee at the authors' institution and was judged to be exempt from oversight. Written informed consent was obtained from all participants entering the National Colorectal Cancer Screening Program.

Study population

Patients between the ages of 50 years and 66 years (40 years and 66 years in case of positive family history of cancer of any type) were advised by their family or general practitioners to participate in the screening. Exclusion criteria were clinical suspicion of colorectal cancer; characteristics that met the criteria for Lynch syndrome, familial adenomatous polyposis, or inflammatory bowel disease; and colonoscopy within the preceding 10 years. For this study, we excluded patients who had screening-detected polyps 10 mm or larger that were not removed (hence histology was unavailable) and patients who had not fully completed the epidemiological questionnaire.

Study procedures and definitions

Patients eligible for screening were asked to complete the epidemiological questionnaire regarding the following potential risk factors for advanced colorectal neoplasia: age,^{9–11} sex,^{10–12} weight and height (to calculate Body Mass Index),^{11–16} family history of colorectal cancer in first-degree relatives,^{10–13} diabetes mellitus,¹⁷ smoking history (number of years of smoking, number of cigarettes smoked per day, and current smoking status),^{14–15} and regular aspirin use (use for at least 3 months at any dose).^{15–18} Information about physical activity, diet, other than aspirin non-steroid anti-inflammatory drug use and alcohol consumption were not collected.

Screening colonoscopy procedures have been described elsewhere.¹⁰ All screening colonoscopists and histopathologists participated in the quality assurance programme.²⁰ Colorectal findings were categorised on the basis of the most advanced lesion identified at screening (including additional required colonoscopies to remove all polyps, when indicated).^{9–10} Advanced neoplasia was defined as cancer or adenoma that was at least 10 mm in diameter, had high-grade dysplasia, had villous or tubulovillous histologic characteristics, or any combination thereof.^{9–10} For the purpose of the analysis, traditional serrated adenomas, sessile serrated lesions, and mixed serrated polyps were categorised as tubular adenomas. Polyps <10 mm in size that were not removed or retrieved were categorised as non-neoplastic.

The following predefined categories of variables were used to analyse the risk factors for detecting advanced neoplasia: age (40–49, 50–54, 55–59, or 60–66 years), sex, family history of colorectal cancer (none, one first-degree relative ≥ 60 years of age with colorectal cancer, one first-degree relative <60 years

of age with colorectal cancer, or two first-degree relatives with colorectal cancer), pack-years smoked (none, <10, 10–19, or ≥ 20 pack-years), diabetes mellitus (yes or no), Body Mass Index (<25, 25–29, or ≥ 30 kg/m²), and regular aspirin use (yes or no). We performed a sensitivity analysis using age as a continuous variable and compared its discriminatory power with the model using age as a categorised variable.

Statistical analysis

The original dataset was randomly partitioned in a 1:1 ratio to generate a test set and a validation set, while controlling for the distribution of the most advanced lesions.²¹ A multivariate logistic regression model was used to investigate the relation between clinical variables and the presence of advanced neoplasia in the test set.²² The likelihood ratio test was used to determine a significant association of a particular variable with the presence of advanced neoplasia and the interaction between variables. For statistically significant effects, the OR and 95% CI were reported for each predefined category of variables. The model was internally validated using the validation set. The Hosmer–Lemeshow test was used to check the goodness-of-fit of the models.²²

The calibration of the model was assessed using the validation set by comparing the expected and observed numbers of patients with advanced neoplasia, overall and for each category of variables.²³ Homogeneous participant groups were defined by all combinations of categories of significant predictors. The expected number of patients with advanced neoplasia for each homogeneous group of study participants was calculated by summing the estimated individual absolute risk predicted by the model developed on the test set. The 95% CIs for the expected to observed ratio were calculated by using normal approximations to Poisson distributions.

The concordance statistic was used to measure models' discrimination among patients with and without advanced neoplasia. For binary logistic regression models, the concordance statistic is equivalent to the area under the receiver-operating characteristic curve.²⁴

The results of the multivariate logistic regression model were used to develop a risk score for detecting advanced neoplasia in asymptomatic patients. Model-adjusted coefficients were rounded up to the nearest one-half integer and then multiplied by two to avoid decimals.¹¹ The performance of the risk score was assessed in the validation set using the concordance statistic.

A p value <0.05 was considered statistically significant. All reported p values are two-sided and not adjusted for multiple testing. The analyses were performed using Stata Statistical Software, V.10 (Stata Corporation, College Station, Texas, USA).

RESULTS

Study participants

Of the 39 265 patients who met the eligibility criteria and underwent colonoscopy in one of the 73 screening centres between January and December 2007, 3347 (8.5%) were excluded, due to incomplete questionnaire feedback (3242 screened participants, 8.3%) or polyps measuring ≥ 10 mm that were not removed (105 screened participants, 0.3%).

The remaining 35 918 patients, 22 164 women (61.7%) and 13 754 men (38.3%), all Caucasians, had a mean age of 55.6 \pm 5.2 years. Of the 35 918 patients, 6897 (19.2%) had a family history of colorectal cancer, 15 678 (43.7%) had history of smoking, 1440 (4.0%) had diabetes mellitus, 7931 (22.1%) had a Body Mass Index ≥ 35 kg/m², and 4623 (12.9%) reported

regular use of aspirin. The characteristics of the study population are summarised in table 1.

Colonoscopy results

Colonoscopy was completed to the caecum in 34 469 patients (96.0%). A total of 6909 patients (19.2%) had an adenoma or cancer. A total of 232 patients (0.6%) had polyps <10 mm in size that were not removed or retrieved, hence were categorised as non-neoplastic abnormalities. Advanced neoplasia was detected in 2544 patients (7.1%), including 336 participants (0.9%) with adenocarcinoma (table 1). Clinically significant complications requiring medical intervention occurred in 42 patients (0.1%) and included seven cases of perforation (three of which occurred after polypectomy), 21 episodes of bleeding, nine cardiovascular events, and five other events. No deaths

occurred as a result of screening colonoscopy or its complications.

Model for the detection of advanced neoplasia

The test and validation sets consisted of 17 979 and 17 939 patients, respectively. We built the multivariate logistic regression model and used the test set to investigate the predictors of detecting advanced neoplasia. The results of the likelihood ratio test indicated significant association between the risk of detecting advanced neoplasia and the following variables: age, sex, family history of colorectal cancer, cigarette smoking and Body Mass Index (table 2). It also revealed significant association of the interaction between sex and Body Mass Index and the risk of detecting advanced neoplasia. The following insignificant variables were reduced from the model: diabetes and regular aspirin use (likelihood ratio test, *p* values equalling 0.24 and

Table 1 Demographic and colonoscopy characteristics for the 35 918 study participants*

Characteristic	All (N=35 918)	Women (N=22 164)	Men (N=13 754)
Age, years			
Range	40–66	40–66	40–66
Mean (SD)	55.6 (5.2)	55.6 (5.1)	55.6 (5.3)
Age group, n (%)			
40–49	3606 (10.0)	2125 (9.6)	1481 (10.8)
50–54	11 000 (30.6)	6818 (30.8)	4182 (30.4)
55–59	12 379 (34.5)	7788 (35.1)	4591 (33.4)
60–66	8933 (24.9)	5433 (24.5)	3500 (25.4)
Family history of CRC, n (%)			
None	29 021 (80.8)	17 846 (80.5)	11 175 (81.2)
One first-degree relative ≥60 years of age with CRC	4798 (13.4)	3000 (13.5)	1798 (13.1)
One first-degree relative <60 years of age with CRC	1809 (5.0)	1121 (5.1)	688 (5.0)
Two first-degree relatives with CRC	290 (0.8)	197 (0.9)	93 (0.7)
Smoking history, pack-year†, n (%)			
None	20 240 (56.4)	13 737 (62.0)	6503 (47.3)
<10	4755 (13.2)	3291 (14.8)	1464 (10.6)
10–19	4576 (12.7)	2663 (12.0)	1913 (13.9)
≥20	6347 (17.7)	2473 (11.2)	3874 (28.2)
Diabetes mellitus, n (%)			
No	34 478 (96.0)	21 404 (96.6)	13 074 (95.1)
Yes	1440 (4.0)	760 (3.4)	680 (4.9)
BMI (kg/m ²), n (%)			
<25	11 439 (31.9)	8258 (37.3)	3181 (23.1)
25–29	16 548 (46.1)	9190 (41.5)	7358 (53.5)
≥30	7931 (22.1)	4716 (21.3)	3215 (23.4)
Regular aspirin use, n (%)			
No	28 960 (80.6)	18 110 (81.7)	10 850 (78.9)
Yes	4623 (12.9)	2590 (11.7)	2033 (14.8)
Not available	2335 (6.5)	1464 (6.6)	871 (6.3)
Total colonoscopy, n (%)	34 469 (96.0)	21 078 (95.1)	13 391 (97.4)
Adequate bowel preparation‡, n (%)	33 909 (94.4)	20 942 (94.5)	12 967 (94.3)
Intravenous sedation, n (%)	20 850 (58.1)	13 565 (61.2)	7285 (53.7)
Main colonoscopy findings, n (%)			
None	14 415 (40.1)	9601 (43.3)	4814 (35.0)
Non-neoplastic abnormalities	14 594 (40.6)	9171 (41.4)	5423 (39.4)
Non-advanced neoplasia§	4365 (12.2)	2207 (10.0)	2158 (15.7)
Advanced neoplasia¶	2544 (7.1)	1185 (5.3)	1359 (9.9)

*Because of rounding, percentages may not total 100.

†Regardless of current smoking status.

‡Bowel preparation was assessed by endoscopists.

§Includes 232 patients with polyps <10 mm in size that were not removed or retrieved.

¶Advanced neoplasia was defined as a cancer or adenoma that was ≥10 mm in diameter, had high-grade dysplasia, had a villous component, or any combination thereof. BMI, Body Mass Index; CRC, colorectal cancer.

Table 2 Associations between individual characteristics and advanced colorectal neoplasia in a multivariable analysis and development of the score (test set, N=17 979)

Covariates	LR test*	Adjusted OR (95% CI)	p Value	Adjusted β coefficient†	Risk score‡
Age group, years	<0.001				
40–49		1	0.002	0	0
50–54		1.53 (1.16 to 2.02)	<0.001	0.43	1
55–59		2.29 (1.75 to 3.00)	<0.001	0.83	2
60–66		3.14 (2.40 to 4.12)		1.15	3
Family history of CRC	<0.001				
None		1	<0.001	0	0
One first-degree relative with CRC, aged ≥ 60 years		1.40 (1.17 to 1.67)	<0.001	0.34	1
One first-degree relative with CRC, aged <60 years		1.66 (1.28 to 2.17)	0.008	0.51	2
Two first-degree relatives with CRC		2.11 (1.22 to 3.66)		0.75	2
Sex	<0.001				
Female		1	<0.001	0	0
Male		2.14 (1.71–2.67)		0.76	2
Smoking history, pack-years§	<0.001				
None		1	0.654	0	0
<10		1.05 (0.86 to 1.27)	0.001	0	0
10–19		1.34 (1.12 to 1.59)	<0.001	0.30	1
≥ 20		1.60 (1.39 to 1.85)		0.47	1
BMI (kg/m ²)	0.033				
<25		1	0.192	0	0
25–29		1.14 (0.94 to 1.38)	0.008	0	0
≥ 30		1.34 (1.08 to 1.67)	0.081	0.30	1 (0)¶
Interaction with male sex**					
25–29		0.78 (0.59 to 1.03)	0.031	0	
≥ 30		0.70 (0.51 to 0.97)		–0.35	

*LR test, the likelihood ratio test for models with and without the specified covariate.

†Non-significant covariates were assigned β coefficient=0.

‡Model-adjusted coefficients were rounded up to the nearest one-half integer and then multiplied by two to avoid decimals.

§Regardless of current smoking status.

¶1 point for female sex and 0 point for male sex.

**Significant interaction was shown only between male sex and BMI.

BMI, Body Mass Index; CRC, colorectal cancer; LR, likelihood ratio test.

0.95, respectively). Table 2 depicts the ORs and 95% CI for each category of a significant variable. Tests for goodness-of-fit of the models in the test and validation datasets permitted acceptance of the fit (p values equalling 0.74 and 0.16, respectively).

The results of the model calibration performed in the validation dataset are shown in table 3. The ratio of expected to observed risk of advanced neoplasia was 1.00 (95% CI 0.95 to 1.06) overall, 1.03 (95% CI 0.97 to 1.12) in women, and 0.98 (95% CI 0.91 to 1.06) in men, indicating good calibration. The concordance statistics of the model were 0.64 for the test set and 0.62 for the validation set, indicating moderate discrimination. A sensitivity analysis performed in a test set, with age as continuous variable showed comparable concordance statistics of the model (0.64, 95% CI 0.63 to 0.66; χ^2 p value=0.82).

The score to predict detection of advanced colorectal neoplasia

The adjusted β coefficients of the logistic regression model fitted on the test set were used to develop the risk score by estimating the likelihood of detecting advanced neoplasia for each category of significant factors (see table 2). The scores for Body Mass Index ≥ 30 kg/m² for different sexes were adjusted according to the interaction coefficient. The score calculated for each person from the validation set estimated the likelihood of detecting advanced neoplasia from 1.32% for patients with a score of 0 to

19.12% for patients with scores of seven and eight (figure 1). The performance characteristics of the score in the validation set are shown in table 4. The concordance statistic for the simplified score in the validation set was 0.62 (95% CI 0.60 to 0.64); the course of the receiver operating characteristics curve is shown in online supplementary figure S1. Online supplementary table S1 depicts the ratio of expected to observed risk for advanced colorectal neoplasia in the validation set by simplified score.

DISCUSSION

Our previous study found that male sex, age of 50 years or more and family history of colorectal cancer were independent risk factors for detecting advanced colorectal neoplasia.¹⁰ In the ensuing discussion, it has been suggested that the observed disparity of advanced neoplasia risk between men and women might have merely reflected sex-based differences in smoking patterns.²⁵ In the present study, we used a new dataset to derive and validate a model for the detection of advanced colorectal neoplasia that included smoking status and other potential confounders, such as age, sex, family history of colorectal cancer, and Body Mass Index. We confirmed previously identified associations, and also found that smoking ≥ 10 pack-years, and Body Mass Index ≥ 30 kg/m² were independent risk factors for detecting advanced colorectal neoplasia. Our study corroborated previously identified risk factors for advanced colorectal neoplasia;^{10–16} it also, for the first time, combined all five

Table 3 Ratio of expected to observed risk for advanced colorectal neoplasia in the validation dataset (N=17 939)

Variables	Observed risk	Expected risk	Expected to observed ratio	(95% CI)
Overall	1270	1275.3	1.00	0.95 to 1.06
Sex				
Female	582	601.5	1.03	0.95 to 1.12
Male	688	673.8	0.98	0.91 to 1.06
Age group, years				
40–49	83	76.2	0.92	0.74 to 1.14
50–54	331	280.1	0.85	0.76 to 0.94
55–59	443	468.6	1.06	0.96 to 1.16
60–66	413	450.4	1.09	0.99 to 1.20
Family history of CRC				
None	1028	1010.2	0.98	0.92 to 1.05
One first-degree relative with CRC, aged ≥60 years	170	176.9	1.04	0.90 to 1.21
One first-degree relative with CRC, aged <60 years	57	69.6	1.22	0.94 to 1.58
Two first-degree relatives with CRC	15	18.6	1.24	0.75 to 2.06
Smoking history, pack-years*				
None	582	609.5	1.05	0.97 to 1.14
<10	164	141.9	0.87	0.74 to 1.01
10–19	200	185.7	0.93	0.81 to 1.07
≥20	324	338.2	1.04	0.94 to 1.16
Women BMI, kg/m ²				
<25	219	193.3	0.88	0.77 to 1.01
25–29	234	256.5	1.10	0.96 to 1.25
≥30	129	151.6	1.18	0.99 to 1.40
Men BMI, kg/m ²				
<25	161	158.7	0.99	0.85 to 1.15
25–29	331	344.1	1.04	0.93 to 1.16
≥30	196	171.0	0.87	0.76 to 1.00

*Regardless of current smoking status.

BMI, Body Mass Index; CRC, colorectal cancer.

important factors and their categories in a multivariate analysis and confirmed obtained results in a validation set.

The present model was well calibrated overall, as well as in men and women, as verified in the validation set, which means that the observed risk of advanced colorectal neoplasia well fitted the expected risk. Therefore, we used the model to develop a simple score for the detection of advanced colorectal

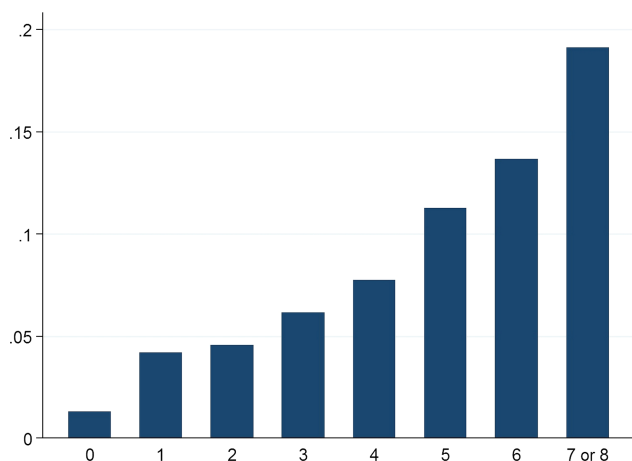


Figure 1 The prevalence of advanced colorectal neoplasia or cancer in the complete dataset (N=35 918) by simplified score. Advanced neoplasia includes advanced adenomas and cancers. Due to small sample sizes, risk scores 7 and 8 are presented together.

neoplasia in asymptomatic patients. The score, based on age, sex, family history of colorectal cancer, smoking status and Body Mass Index, estimated the likelihood of detecting advanced colorectal neoplasia in the validation set from 1.32% to 19.12% in patients with 0 to 7–8 points, respectively. The estimation of individual risk of detecting advanced colorectal neoplasia may help asymptomatic patients and healthcare providers to make informed decisions about screening.²⁶ For example, the likelihood of detecting advanced colorectal neoplasia in a 53-year-old, overweight, never-smoking woman, with one first-degree relative 60 years of age or older with colorectal cancer, is difficult to compare with that of a 56-year-old man who smoked for 20 pack-years, but has healthy weight and no family history of colorectal cancer. However, based on the results of the present model, the respective likelihood of detecting advanced neoplasia for two such patients are 4.65% and 12.46% (or 4.57% and 11.27%, respectively, using simplified scoring). Such results do not mean that one should discourage the woman from participation in an existing screening programme in a given country aiming at average risk group; rather they indicate that the man should be specifically encouraged to be screened, because the likelihood of detecting advanced neoplasia in his colorectum is almost twice that of the average screening population. For ease of clinical application, the present model could be transformed into an online calculator of the likelihood of detecting advanced colorectal neoplasia and used in mobile easy access media. Although lack of symptoms and low perceived risk of colorectal cancer are considered major barriers to screening,^{7 8} it is unknown, whether providing the

Table 4 Performance characteristics of the score in the validation set (N=17 939)

Risk score	Persons, n	Advanced neoplasia*					
		N	Per cent	Sensitivity, %	Specificity, %	Positive predictive value, %	Negative predictive value, %
0	152	2	1.32	100.00	0.00	7.08	–
1	2259	95	4.21	99.84	0.90	7.13	98.68
2	3608	165	4.57	92.36	13.88	7.55	95.98
3	4626	284	6.14	79.37	34.54	8.46	95.65
4	3745	290	7.74	57.01	60.59	9.93	94.87
5	2432	274	11.27	34.17	81.31	12.23	94.19
6	981	134	13.66	12.60	94.26	14.32	93.40
7 and 8†	136	26	19.12	2.05	99.34	19.12	93.01
Total	17 939	1270	7.08				

*Advanced neoplasia includes advanced adenomas and cancers.

†Due to small sample risk scores 7 and 8 are presented in one row.

estimate of individual risk could facilitate the informed decision to undergo endoscopy screening in a similar way as it worked for prostate cancer screening.²⁷ It is particularly unknown, what kind of effect on participation in screening, would have the lower than average estimate of likelihood of detecting advanced colorectal neoplasia.

Another potential application of a model for the detection of advanced colorectal neoplasia is to guide practical recommendations for mass screening; however, this application would require a model with high discriminatory power.²⁸ The present model had only moderate concordance statistic value, comparable to that of previously published models for the detection of advanced colorectal neoplasia in Western populations,^{11 29} even though the present model included more risk factors than previous models did. Three issues may explain this observation. First, the model of Betes *et al*¹¹ lacked validation, which may have led to overestimation of its discriminatory power. Second, the models of Betes *et al*¹¹ and Lin *et al*²⁹ were derived from populations with a broader age range, which may have increased their discriminatory power, because age is the most powerful clinical risk factor for advanced colorectal neoplasia. Third, additional independent risk factors included in the present model were too weak to significantly change its discriminatory power.³⁰ The models for the detection of advanced colorectal neoplasia in East Asian populations demonstrated variable discriminatory power.^{31 32} The model by Yeoh *et al*³¹ demonstrated discriminatory power comparable to that achieved in Western populations, while the model by Cai QC *et al*³² demonstrated better discrimination. The latter model missed family history of colorectal cancer but included various dietary factors, which (in contrast to previous studies^{15 16}) showed strong association with the risk of advanced colorectal neoplasia; however, these factors are prone to recall bias. Therefore, it is rather unlikely that a model based on simple, reliable clinical factors alone would ever have sufficient discriminatory power to limit the target population for screening. It is corroborated by the results of a very recent study by Tao S *et al*.³³ The model that included risk factors missing in the present study (alcohol consumption, red meat consumption, ever regular use of non-steroidal anti-inflammatory drugs, previous colonoscopy and previous detection of polyps), also demonstrated moderate discriminatory value.³³ On the other hand, indirect comparison suggests that the present model's sensitivity and specificity for advanced colorectal neoplasia may be comparable with a single round guaiac faecal occult blood test (mostly due to poor diagnostic performance of guaiac faecal occult blood test for

detection of advanced adenomas).^{34 35} Although the model has considerably lower discriminatory power for advanced colorectal neoplasia compared to the one reported for faecal immunochemical tests,^{34–37} it has been shown that combining clinical risk factors with faecal immunochemical test outcome results in improved discrimination.^{38 39} Therefore, it is likely that in the future the clinical factors identified in the model will be combined with results of faecal immunochemical test and/or blood-based biomarkers to select a target population for colonoscopy.

Our study has certain notable features. Despite a large sample size, we have not identified any statistically significant association between diabetes mellitus or aspirin use and the risk of advanced colorectal neoplasia. Although diabetes mellitus is a known risk factor for colorectal cancer,¹⁷ its association with advanced colorectal neoplasia is less certain.³¹ The observed lack of association may also be due to the lower-than-expected prevalence of diabetes mellitus in the study cohort,⁴⁰ which likely reflects recall bias or self-selection to opportunistic screening.⁴¹

The lack of a statistically significant association between aspirin use and the risk of advanced colorectal neoplasia in our study may be due to recall bias or missing data regarding the dose and regularity of aspirin use. Moreover, we have not collected the data on other non-steroid anti-inflammatory drugs, which in some studies were analysed together with aspirin.¹⁵

Advanced neoplasia, not just cancer, was chosen for analysis because it has been suggested as the most appropriate target for endoscopy screening.^{9–12 15 29 31 32} Although some previous risk prediction models were developed for cancer alone,^{23 42 43} cancers and advanced neoplasia are surrogate endpoints of primary cancer screening endpoint, which is colorectal cancer mortality.⁴⁴ Early detection and treatment of colorectal cancer is associated with a reduction in colorectal cancer mortality,² but a detection and removal of adenomas, especially advanced ones, is associated with additional reduction in colorectal cancer incidence and mortality.^{45 46} Therefore, it is uncertain, whether cancer alone or advanced neoplasia is a better endpoint for risk prediction models, but the latter may be particularly suited for use in endoscopy screening.

The primary endpoint of our model was advanced neoplasia located anywhere in the colorectum, therefore, the risk score is not optimised for sigmoidoscopy screening. Nevertheless, we built an additional model to investigate risk factors for detecting distal advanced neoplasia, using sigmoid-descending colon junction as an artificial boundary between distal and proximal colon. The model for the detection of distal advanced neoplasia

identified the same risk factors and showed comparable discriminatory power (data not shown).

The limitations of our study require comment. First, the validation process was limited because it was performed in a dataset that was randomly selected from a population recruited in the same setting. The model's performance has not been tested outside Poland or in non-Caucasians. Nonetheless, the National Colorectal Cancer Screening Program recruited participants in 73 centres located in all administrative and geographic regions of Poland, and was open free of charge to all eligible Polish citizens, providing our study with sociodemographic diversity. Moreover, the prevalence of advanced colorectal neoplasia identified in our study was 7.1%, which is within the range of values reported in studies performed in the USA^{9 15 29} and Europe.^{5 10 11} Additionally, the adjusted ORs for detecting advanced colorectal neoplasia in various categories of risk factors in our study are similar to that reported in previously published large studies.^{10 31 32 47 48} Notably, in our previous study,¹⁰ performed several years before and in different endoscopy centres, we used the same key to categorise age, family history of colorectal cancer and gender and yielded virtually the same adjusted ORs for each category of variables.

Second, our cohort does not fully cover the recommended age range for screening (people aged 67–75 years were not included) what may limit the applicability of the results for the entire population eligible for screening.⁴⁹ On the other hand, by including people at the lower age range for screening (people aged 40–49 years with family history of cancer), this model may help to identify and encourage younger people at considerable risk to undergo screening.

Third, given the cross-sectional design of the study, our risk score is suitable only to predict the detection of advanced neoplasia at the present time, and not the future risk of developing advanced colorectal neoplasia or dying from colorectal cancer.

In summary, we derived and internally validated a model that predicts the likelihood of detecting advanced colorectal neoplasia in asymptomatic Caucasian patients based on age, sex, smoking habits, Body Mass Index, and family history of colorectal cancer. The results of the model were used to develop a simple score that estimates the likelihood of detecting advanced colorectal neoplasia. Once externally validated, the score may be useful for counselling or designing primary prevention studies.

Acknowledgements The authors are grateful to all the endoscopists and histopathologists who participated in the National Colorectal Cancer Screening Program in Poland. The authors thank Michael Bretthauer, MD, PhD, and Øyvind Holme, MD, from the Department of Management and Health Economy University of Oslo, and the Department of Transplantation Medicine, Gastroenterology, Oslo University Hospital at Rikshospitalet, Oslo, Norway, for their editorial review of an earlier draft of this article. The authors thank Tomasz Burzykowski, PhD, from the Interuniversity Institute for Biostatistics and Statistical Bioinformatics, Hasselt University, Diepenbeek, Belgium, Mitchell Gail, MD, PhD, from the Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, and Krzysztof Przewozniak, MSc, from the Cancer Epidemiology and Prevention Department, the Maria Skłodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw, Poland, for their advice.

Contributors The study was designed by the first three investigators, and the first author wrote the first draft of the manuscript. All authors participated in data collection and analysis, contributed to the manuscript, approved the final version of the manuscript, agreed to submit the manuscript for publication, and vouched for the completeness and accuracy of the data.

Funding This study was supported by the Polish Ministry of Health and the Polish Foundation of Gastroenterology. Michal F Kaminski received a stipend from the Foundation for Polish Science during the study period.

Competing interests None.

Patient consent Obtained.

Ethics approval the Research Ethical Committee at the Maria Skłodowska-Curie Memorial Cancer Centre and Institute of Oncology.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>

REFERENCES

- Littlejohn C, Hilton S, Macfarlane GJ, *et al*. Systematic review and meta-analysis of the evidence for flexible sigmoidoscopy as a screening method for the prevention of colorectal cancer. *Br J Surg* 2012;99:1488–500.
- Hewitson P, Glasziou P, Irwig L, *et al*. Screening for colorectal cancer using the faecal occult blood test, Hemoccult. *Cochrane Database Syst Rev* 2007;(1): CD001216.
- von Karsa L, Patnick J, Segnan N, *et al*. European guidelines for quality assurance in colorectal cancer screening and diagnosis: overview and introduction to the full supplement publication. *Endoscopy* 2013;45:51–9.
- Logan RF, Patnick J, Nickerson C, *et al*. Outcomes of the Bowel Cancer Screening Programme (BCSP) in England after the first 1 million tests. *Gut* 2012;61:1439–46.
- Pox CP, Altenhofen L, Brenner H, *et al*. Efficacy of a nationwide screening colonoscopy program for colorectal cancer. *Gastroenterology* 2012;142:1460–7 e2.
- Khalid-de Bakker C, Jonkers D, Smits K, *et al*. Participation in colorectal cancer screening trials after first-time invitation: a systematic review. *Endoscopy* 2011;43:1059–86.
- Vernon SW. Participation in colorectal cancer screening: a review. *J Natl Cancer Inst* 1997;89:1406–22.
- Klabunde CN, Vernon SW, Nadel MR, *et al*. Barriers to colorectal cancer screening: a comparison of reports from primary care physicians and average-risk adults. *Med Care* 2005;43:939–44.
- Lieberman DA, Weiss DG, Bond JH. Use of colonoscopy to screen asymptomatic adults for colorectal cancer. Veterans Affairs Cooperative Study Group 380. *N Engl J Med* 2000;343:162–68.
- 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.
- 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.
- Nguyen SP, Bent S, Chen YH, *et al*. Gender as a risk factor for advanced neoplasia and colorectal cancer: a systematic review and meta-analysis. *Clin Gastroenterol Hepatol* 2009;7:676–81 e1–3.
- Taylor DP, Burt RW, Williams MS, *et al*. Population-based family history-specific risks for colorectal cancer: a constellation approach. *Gastroenterology* 2010;138:877–85.
- Botteri E, Iodice S, Raimondi S, *et al*. Cigarette smoking and adenomatous polyps: a meta-analysis. *Gastroenterology* 2008;134:388–95.
- Lieberman DA, Prindiville S, Weiss DG, *et al*. Risk factors for advanced colonic neoplasia and hyperplastic polyps in asymptomatic individuals. *JAMA* 2003;290:2959–67.
- Terry MB, Neugut AI, Bostick RM, *et al*. Risk factors for advanced colorectal adenomas: a pooled analysis. *Cancer Epidemiol Biomarkers Prev* 2002;11:622–9.
- Larsson SC, Orsini N, Wolk A. Diabetes mellitus and risk of colorectal cancer: a meta-analysis. *J Natl Cancer Inst* 2005;97:1679–87.
- Cole BF, Logan RF, Halabi S, *et al*. Aspirin for the chemoprevention of colorectal adenomas: meta-analysis of the randomized trials. *J Natl Cancer Inst* 2009;101:256–66.
- Larsson SC, Wolk A. Meat consumption and risk of colorectal cancer: a meta-analysis of prospective studies. *Int J Cancer* 2006;119:2657–64.
- Kaminski MF, Regula J, Kraszewska E, *et al*. Quality indicators for colonoscopy and the risk of interval cancer. *N Engl J Med* 2010;362:1795–803.
- Neter J, Kutner MH, Nachtsheim CJ, *et al*. *Applied linear statistical models*. 4th edn. Chicago: Irwin, 1996.
- Hosmer DWJ, Lemeshow S. *Applied logistic regression*. New York: John Wiley, 1989.
- 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.
- Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29–36.
- Latreille MW, Anderson JC. Colonoscopy screening for detection of advanced neoplasia. *N Engl J Med* 2007;356:632; author reply 33–4.
- Foster P, Anderson CM. Reaching targets in the national cervical screening programme: are current practices unethical? *J Med Ethics* 1998;24:151–7.

- 27 van Vugt HA, Roobol MJ, Venderbos LD, *et al.* Informed decision making on PSA testing for the detection of prostate cancer: an evaluation of a leaflet with risk indicator. *Eur J Cancer* 2010;46:669–77.
- 28 Gail MH, Pfeiffer RM. On criteria for evaluating models of absolute risk. *Biostatistics* 2005;6:227–39.
- 29 Lin OS, Kozarek RA, Schembre DB, *et al.* Risk stratification for colon neoplasia: screening strategies using colonoscopy and computerized tomographic colonography. *Gastroenterology* 2006;131:1011–19.
- 30 Pepe MS, Janes H, Longton G, *et al.* Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol* 2004;159:882–90.
- 31 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.
- 32 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.
- 33 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* 2013. Published Online First: 8 Sept 2013. pii: S1542-3565(13)01297-4. doi: 10.1016/j.cgh.2013.08.042
- 34 Park DI, Ryu S, Kim YH, *et al.* Comparison of guaiac-based and quantitative immunochemical fecal occult blood testing in a population at average risk undergoing colorectal cancer screening. *Am J Gastroenterol* 2010;105:2017–25.
- 35 Brenner H, Tao S. Superior diagnostic performance of faecal immunochemical tests for haemoglobin in a head-to-head comparison with guaiac based faecal occult blood test among 2235 participants of screening colonoscopy. *Eur J Cancer* 2013;49:3049–54.
- 36 Haug U, Hundt S, Brenner H. Quantitative immunochemical fecal occult blood testing for colorectal adenoma detection: evaluation in the target population of screening and comparison with qualitative tests. *Am J Gastroenterol* 2010;105:682–90.
- 37 Hundt S, Haug U, Brenner H. Comparative evaluation of immunochemical fecal occult blood tests for colorectal adenoma detection. *Ann Intern Med* 2009;150:162–9.
- 38 Stegeman I, de Wijkerslooth TR, Stoop EM, *et al.* Combining risk factors with faecal immunochemical test outcome for selecting CRC screenees for colonoscopy. *Gut* 2014;63:466–71.
- 39 Omata F, Shintani A, Isozaki M, *et al.* Diagnostic performance of quantitative fecal immunochemical test and multivariate prediction model for colorectal neoplasms in asymptomatic individuals. *Eur J Gastroenterol Hepatol* 2011;23:1036–41.
- 40 Polakowska M, Piotrowski W. Incidence of diabetes in the Polish population: results of the Multicenter Polish Population Health Status Study—WOBASZ. *Pol Arch Med Wewn* 2011;121:156–63.
- 41 Wiszniewski M, Kaminski MF, Kraszewska E, *et al.* Distribution of life style factors and family history of colorectal cancer and the risk of advanced neoplasia in opportunistic versus organized screening. *Gastroenterology* 2011;140:S412.
- 42 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.
- 43 Freedman AN, Slattery ML, Ballard-Barbash R, *et al.* Colorectal cancer risk prediction tool for white men and women without known susceptibility. *Am J Clin Oncol* 2009;27:686–93.
- 44 Cuzick J, Cafferty FH, Edwards R, *et al.* Surrogate endpoints for cancer screening trials: general principles and an illustration using the UK Flexible Sigmoidoscopy Screening Trial. *J Med Screen* 2007;14:178–85.
- 45 Winawer SJ, Zauber AG, Ho MN, *et al.* Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. *N Engl J Med* 1993;329:1977–81.
- 46 Zauber AG, Winawer SJ, O'Brien MJ, *et al.* Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. *N Engl J Med* 2012;366:687–96.
- 47 Stegeman I, de Wijkerslooth TR, Stoop EM, *et al.* Colorectal cancer risk factors in the detection of advanced adenoma and colorectal cancer. *Cancer Epidemiol* 2013;37:278–83.
- 48 Kolligs FT, Crispin A, Munte A, *et al.* Risk of advanced colorectal neoplasia according to age and gender. *PLoS ONE* 2011;6:e20076.
- 49 Lansdorp-Vogelaar I, von Karsa L. European guidelines for quality assurance in colorectal cancer screening and diagnosis. First Edition—Introduction. *Endoscopy* 2012;44(Suppl 3):SE15–30.