ORIGINAL RESEARCH—CLINICAL

Development and External Validation of a Prediction Model for Colorectal Cancer Among Patients Awaiting Surveillance Colonoscopy Following Polypectomy

Theodore R. Levin,[1,2,*] Christopher D. Jensen,[1,*] Amy R. Marks,[1]
David Schlessinger,[1] Vincent Liu,[1] Natalia Udaltsova,[1] Jessica Badalov,[1]
Evan Layefsky,[1] Douglas A. Corley,[1] Joshua R. Nugent,[1] and Jeffrey K. Lee[1]

[1]Division of Research, Kaiser Permanente Northern California, Oakland, California; and [2]Gastroenterology Department, Kaiser Permanente Medical Center, Walnut Creek, California

BACKGROUND AND AIMS: Demand for surveillance colonoscopy can sometimes exceed capacity, such as during and following the coronavirus disease 2019 pandemic, yet no tools exist to prioritize the patients most likely to be diagnosed with colorectal cancer (CRC) among those awaiting surveillance colonoscopy. We developed a multivariable prediction model for CRC at surveillance comparing performance to a model that assigned patients as low or high risk based solely on polyp characteristics (guideline-based model). **METHODS:** Logistic regression was used for model development among patients receiving surveillance colonoscopy in 2014–2019. Candidate predictors included index colonoscopy indication, findings, and endoscopist adenoma detection rate, and patient and clinical characteristics at surveillance. Patients were randomly divided into model development (n = 36,994) and internal validation cohorts (n = 15,854). External validation was performed on 30,015 patients receiving surveillance colonoscopy in 2020–2022, and the multivariable model was then updated and retested. **RESULTS:** One hundred fourteen, 43, and 71 CRCs were detected at surveillance in the 3 cohorts, respectively. Polyp size ≥10 mm, adenoma detection rate <32.5% or missing, patient age, and ever smoked tobacco were significant CRC predictors; this multivariable model outperformed the guideline-based model (internal validation cohort area under the receiver-operating characteristic curve: 0.73, 95% confidence interval (CI): 0.66–0.81 vs 0.52, 95% CI: 0.45–0.60). Performance declined at external validation but recovered with model updating (operating characteristic curve: 0.72 95% CI: 0.66–0.77). **CONCLUSION:** When surveillance colonoscopy demand exceeds capacity, a prediction model featuring common clinical predictors may help prioritize patients at highest risk for CRC among those awaiting surveillance. Also, regular model updates can address model performance drift.

Keywords: Prediction Model; Post Polypectomy; Colonoscopy Surveillance; Colorectal Cancer; Risk Stratification

Introduction

Colonoscopy prevents colorectal cancer (CRC) and reduces mortality through removal of precancerous polyps and detection of cancer at an earlier, more treatable stage.[1] Despite polyp removal, some patients remain at elevated risk for CRC and are recommended to undergo colonoscopy surveillance.[2] Surveillance is a common colonoscopy indication and at times (eg, during and following the coronavirus disease 2019 [COVID-19] pandemic), demand for colonoscopy exceeds capacity, necessitating that patients requiring surveillance colonoscopy be placed on a waiting list.[3–5] However, for providers managing these waiting lists, there are no tools available to prioritize those most likely to be diagnosed with CRC among the patients awaiting surveillance colonoscopy.

To address this need, we sought to develop and validate a multivariable prediction model (*multivariable model*) to identify, among patients awaiting surveillance colonoscopy, those at highest risk for CRC, and to compare performance to a *guideline-based model* where patients were categorized as high or low risk based on their prior polyp findings (ie, number, size, and histology of precancerous polyps removed).[2]

Methods

Study Design, Setting, Funding, and Oversight

This cross-sectional study examined the relationship between candidate predictors and risk of CRC diagnosis at surveillance colonoscopy among members of Kaiser Permanente Northern California (KPNC), an integrated health-care system

serving approximately 4.5 million members across 21 medical centers in urban, suburban, and semirural regions throughout Northern California. KPNC's membership is diverse and similar in socioeconomic characteristics to the region's census demographics, including the proportions with commercial insurance, Medicare, or Medicaid.[6,7] The study was funded by The Permanente Medical Group's Delivery Science and Applied Research Program and was approved by the KPNC institutional review board.

Eligibility Criteria

Model development and internal validation cohorts. Individuals were eligible for inclusion in the prediction model development and internal validation cohorts if they 1) were KPNC health plan members who had a *surveillance* colonoscopy performed in 2014–2019, for an *index* (baseline) colonoscopy with polypectomy performed ≥2 years prior or as surveillance for a prior colonoscopy with polypectomy (this included individuals who had a code for history of adenoma from a procedure performed outside the health system); 2) had ≥1 year of membership (allowing for a ≤2-month gap) before the date of the surveillance colonoscopy; and 3) were 45–85 years of age at the date of the surveillance colonoscopy. Individuals were excluded from the study if, before their surveillance colonoscopy, they were at a substantially modified risk for CRC (ie, prior history of CRC, inflammatory bowel disease, hereditary CRC syndrome, or total colectomy); if they had a colonoscopy <2 years before the surveillance colonoscopy; if the colonoscopy taking place after the index procedure had a diagnostic indication (eg, iron deficiency anemia, positive fecal test indication, etc.) because this would remove them from the surveillance colonoscopy waiting list; or if they had a subsequent hereditary CRC syndrome diagnosis. If a member had more than 1 surveillance colonoscopy performed during the study period, only the first was included (see below for definitions and data sources). Cohorts were comprised of all subjects who met the study eligibility criteria and sample size calculations were not performed.

External validation cohort. The same eligibility criteria as above were used for the external validation cohort except that cohort members had a surveillance colonoscopy performed during a separate study period of 2020–2022.

Outcome

The outcome was CRC, defined as adenocarcinoma of the colon or rectum diagnosed at or within 6 months after the surveillance colonoscopy; due to a maximum diagnosis date of April 30, 2023, surveillance colonoscopies performed in November to December 2022 had a follow-up interval of 4–6 months. While 75% or more of CRCs were diagnosed within 1 week after surveillance colonoscopy in the 3 cohorts, diagnoses made up to 6 months after the procedure date were considered part of the surveillance episode, given that events such as a subsequent surgery or a repeat colonoscopy (eg, due to initially poor bowel preparation) may lead to CRC diagnosis.

Candidate Predictors and Definitions

Candidate CRC risk predictors (Table 1) were identified by literature review and included the following categories.

Demographic and clinical variables, collected at the time of the surveillance colonoscopy, included patient age (per 1-year increase); sex (male or female); race and ethnicity (Asian or Pacific Islander [non-Hispanic], Black [non-Hispanic], Hispanic, White [non-Hispanic], or other and unknown as reported in the electronic health record); first-degree family history of CRC (yes or either no or missing); body mass index (per 1 kg/m2 increase); ever smoked tobacco (yes or either no or missing); and diabetes diagnosis (yes or no). Index colonoscopy procedure-related variables included whether the procedure was performed at KPNC (yes or no); extent of examination (complete to cecum, incomplete, or missing); bowel preparation adequacy (good or excellent, fair or poor, or missing); procedure indication (screening, diagnostic, positive fecal test, surveillance, or missing); adenoma detection rate (ADR) of the endoscopist who performed the index procedure based on all colonoscopy indications (≥32.5%, <32.5%, or missing) for the year before the index colonoscopy; and time interval from index colonoscopy to surveillance colonoscopy (2 to <3, 3 to <5, 5 to <7, 7 to <10, ≥10 years, or missing). Index colonoscopy finding-related variables included any adenoma (yes, no, or missing); adenoma with advanced histology (yes, no, or missing); maximum polyp size (≥10 mm, <10 mm, no polyp, or missing) and polyps ≥10 mm in size were assumed to be adenomas; number of adenomas (0 to 2, ≥3, or missing); worst finding (advanced adenoma, nonadvanced adenoma, no adenoma, or missing); and patient risk status according to current post polypectomy surveillance guidelines (high risk, low risk, or missing).

Adenoma with advanced histology was defined as an adenoma with villous or tubulovillous features, adenocarcinoma in situ, or high-grade dysplasia. *Advanced adenoma* was defined as an adenoma with advanced histology or an adenoma or serrated polyp ≥10 mm. For the variable, patient risk status according to post polypectomy surveillance guidelines,[2] high risk was defined as having an advanced adenoma (ie, adenoma with advanced histology or an adenoma or serrated polyp ≥10 mm) or 3 or more adenomas at the index colonoscopy, while low risk was defined as having 1–2 nonadvanced adenomas <10 mm in size or no adenoma at the most recent prior colonoscopy but with a history of adenoma or serrated polyp on a remote colonoscopy.

All-indication ADRs were calculated for each physician for each calendar year using all colonoscopy examinations performed on patients 45–85 year old, and an ADR value was assigned using the endoscopist's ADR for the calendar year before the index colonoscopy. The ADR was considered missing if the endoscopist was unknown or if a physician performed fewer than 100 total colonoscopies in a calendar year (considered too few for a stable estimate). In analyses, those with missing ADR values were grouped with those with an ADR <32.5% based on the assumption that endoscopists with fewer than 100 total colonoscopies in a year would be more likely to have a lower ADR. Also, for ease of use in clinical settings, we included ADR in the model as a categorical rather than continuous variable, since whether an endoscopist is a high or low adenoma detector is more likely to be known than their actual individual all-indication ADR value in any given year. The median ADR based on all colonoscopy indications was 36% in the development cohort, and therefore we tested ADRs of 30%, 32.5%, 35%, 37.5%, and 40% in this cohort; values ≥ 32.5% yielded the strongest association with CRC risk and was used in

Table 1. Characteristics of the Model Development, Internal Validation, and External Validation Cohorts

Characteristics	Model development cohort 2014–2019	Internal validation cohort 2014–2019	P value	External validation cohort 2020–2022	P value
Cohort members, n	36,994	15,854		38,242	
Patient demographic and clinical characteristics at surveillance					
Age, y					
Mean (SD)	*66 (8)*	*66 (8)*	.96	*67 (8)*	<.001
Sex, n (%)			.29		.29
Female	15,403 (41.6)	6522 (41.1)		15,776 (41.3)	
Male	21,591 (58.4)	9332 (58.9)		22,466 (58.7)	
Race and ethnicity, n (%)			.79		<.001
Asian or Pacific Islander	6079 (16.4)	2570 (16.2)		6973 (18.2)	
Black	2104 (5.7)	875 (5.5)		2299 (6.0)	
Hispanic	3832 (10.4)	1673 (10.6)		4199 (11.0)	
White	18,785 (50.8)	8106 (51.1)		15,437 (40.4)	
Other and missing	6194 (16.7)	2630 (16.6)		9334 (24.4)	
First-degree family history of colorectal cancer, n (%)			.13		.55
Yes	6777 (18.3)	2816 (17.8)		6941 (18.2)	
No (default)	30,217 (81.7)	13,038 (82.2)		31,301 (81.8)	
Body mass index (kg/m^2)					
Mean (SD)	*29 (6)*	*29 (6)*	.94	*28 (6)*	<.001
Median (IQR)	*28 (25–32)*	*28 (25–32)*	.98	*27 (24–31)*	<.001
Ever smoked tobacco, n (%)			.70		<.001
Yes	13,220 (35.7)	5693 (35.9)		11,520 (30.1)	
No (default)	23,774 (64.3)	10,161 (64.1)		26,722 (69.9)	
Diabetes			.35		<.001
Yes	7444 (20.1)	3247 (20.5)		8306 (21.7)	
No	29,550 (79.9)	12,607 (79.5)		29,936 (78.3)	
Index colonoscopy-related characteristics					
Performed within KPNC (and have information), n (%)			.81		<.001
Yes	34,422 (93.0)	14,761 (93.1)		37,275 (97.5)	
No	2572 (7.0)	1093 (6.9)		967 (2.5)	
Extent of examination, n (%)			<.05		<.001
Complete to cecum	33,324 (90.1)	14,311 (90.3)		36,574 (95.6)	
Incomplete	181 (0.5)	53 (0.3)		158 (0.4)	
Missing	3489 (9.4)	1490 (9.4)		1510 (3.9)	
Bowel preparation, n (%)			.64		<.001
Good or excellent	27,937 (75.5)	12,018 (75.8)		33,367 (87.3)	
Poor or fair	3982 (10.8)	1710 (10.8)		2898 (7.6)	
Missing	5075 (13.7)	2126 (13.4)		1977 (5.2)	
Colonoscopy indication, n (%)			.96		<.001
Screening	10,292 (27.8)	4362 (27.5)		9772 (25.6)	
Diagnostic	8336 (22.5)	3593 (22.7)		7484 (19.6)	
Positive fecal test	6987 (18.9)	3008 (19.0)		7615 (19.9)	
Surveillance	8807 (23.8)	3798 (24.0)		12,404 (32.4)	
Missing	2572 (7.0)	1093 (6.9)		967 (2.5)	
ADR, all indications, %			.70		
Median (IQR)	*36 (31–43)*	*36 (31–43)*	*.19*	*46 (39–53)*	*<.001*
Missing, n (%)	4795 (13.0)	2055 (13.0)		2933 (9.8)	
Cutoff 32.5%			.70		<.001
<32.5 or missing	14,722 (39.8)	6338 (40.0)			
≥32.5	22,272 (60.2)	9516 (60.0)			
Cutoff 37.5%					<.001
<37.5 or missing				10,829 (28.3)	
≥37.5				27,413 (71.7)	
Time from index to surveillance colonoscopy, y, n (%)			.91		<.001
2 to <3	1044 (2.8)	453 (2.9)		1624 (4.2)	
3 to <5	9281 (25.1)	4009 (25.3)		15,608 (40.8)	
5 to <7	18,642 (50.4)	8022 (50.6)		13,748 (36.0)	

Table 1. Continued

Characteristics	Model development cohort 2014–2019	Internal validation cohort 2014–2019	P value	External validation cohort 2020–2022	P value
7 to <10	4750 (12.8)	1979 (12.5)		4831 (12.6)	
≥10	705 (1.9)	298 (1.9)		1464 (3.8)	
Missing	2572 (7.0)	1093 (6.9)		967 (2.5)	
Mean (SD)	*6 (2)*	*6 (2)*	*.43*	*5 (2)*	*<.001*
Median (IQR)	*5 (5–6)*	*5 (5–6)*	*.85*	*5 (4–6)*	*<.001*
Index colonoscopy findings					
Any adenoma, n (%)			.83		<.001
Yes	25,296 (68.4)	10,884 (68.6)		29,176 (76.3)	
No	9126 (24.7)	3877 (24.5)		8099 (21.2)	
Missing	2572 (7.0)	1093 (6.9)		967 (2.5)	
Adenoma with advanced histology, n (%)			.73		<.001
Yes	2833 (7.7)	1185 (7.5)		2592 (6.8)	
No	31,598 (85.4)	13,576 (85.6)		34,683 (90.7)	
Missing	2572 (7.0)	1093 (6.9)		967 (2.5)	
Polyp size ≥10 mm, n (%)			.22		<.01
Yes	4146 (11.2)	1697 (10.7)		4061 (10.6)	
No	30,276 (81.8)	13,064 (82.4)		33,214 (86.9)	
Missing	2572 (7.0)	1093 (6.9)		967 (2.5)	
3 or more adenomas, n (%)			.77		<.001
Yes	2958 (8.0)	1296 (8.2)		10,953 (28.6)	
No	31,464 (85.1)	13,465 (84.9)		26,322 (68.8)	
Missing	2572 (7.0)	1093 (6.9)		967 (2.5)	
Worst finding, n (%)			.38		<.001
Advanced adenoma (size ≥10 mm or advanced histology)	5696 (15.4)	2364 (14.9)		5693 (14.9)	
Nonadvanced adenoma	19,618 (53.0)	8525 (53.8)		23,493 (61.4)	
No adenoma	9108 (24.6)	3872 (24.4)		8089 (21.2)	
Missing	2572 (7.0)	1093 (6.9)		967 (2.5)	
Risk status based on postpolypectomy surveillance guidelines, n (%)			.50		<.001
High risk	7608 (20.6)	3193 (20.1)		13,879 (36.3)	
Low risk	26,814 (72.5)	11,568 (73.0)		23,396 (61.2)	
Missing	2572 (7.0)	1093 (6.9)		967 (2.5)	

P values for the medians are italicized.
Characteristics for the internal and external validation cohorts were compared to those for the development cohort. Chi-square tests, 2-sample t-tests, and Wilcoxon rank-sum tests were used to compare proportions, mean, and median values, respectively.
ADR, adenoma detection rate; IQR, interquartile range; KPNC, Kaiser Permanente Northern California; SD, standard deviation.

the multivariable model. The median ADR in the external validation cohort was 46% and 37.5% was selected as the threshold using a similar approach.

Data Sources

Patient demographic and clinical characteristics were ascertained relative to the surveillance colonoscopy date and obtained from KPNC electronic medical records and databases. History of cancer, inflammatory bowel disease, and hereditary CRC syndromes were ascertained from the KPNC cancer registry and International Classification of Diseases 9th and 10th revision codes. History of total colectomy was ascertained from Current Procedural Terminology codes. Colonoscopies were identified using industry-standard codes (ie, Current Procedural Terminology, International Classification of Diseases, Healthcare Common Procedural Coding System) and local procedure codes. Colonoscopy indication was ascertained by a validated colonoscopy indication algorithm based on symptoms and conditions identified using electronic medical records.[8] Colonoscopy quality measures (ie, extent of the examination and bowel preparation quality) were ascertained from discrete data fields and text analysis of colonoscopy reports using both SAS software v9.4 (SAS Instiute Inc., Cary, North Carolina) and a commercial natural language processing software (Linguamatics I2E, www.linguamatics.com; United Kingdom). Colonoscopy findings (ie, polyps, adenomas [including conventional, traditional serrated, and sessile serrated adenomas], villous histology, adenocarcinoma in situ) were identified using Systematized Nomenclature of Medicine (SNOMED) codes. High-grade dysplasia and sessile serrated polyps have no specific SNOMED codes and were identified using text string searches of pathology reports. Polyp size was ascertained from colonoscopy reports using SAS and the natural language processing

software. The number of pathology jars with a conventional adenoma (as determined by SNOMED codes) was used as a surrogate marker for the total number of conventional adenomas per patient. Cancer diagnoses before mid-2021 were obtained from the KPNC cancer registry, which reports to the Surveillance, Epidemiology and End-Results program and captures >98% of cancers diagnosed among members compared with manual review. For CRC diagnosed in 2022–2023, potential cases were flagged using SNOMED codes, International Classification of Diseases 10th revision diagnostic codes, and synoptic reporting of surgical pathology data, and then confirmed by manual chart review. Among 219,060 patients with a colonoscopy in 2018–2019, compared to the cancer registry, this proxy method had a sensitivity of 0.941 (correctly identified 3172 of 3371 CRC cases) and specificity of 0.996 (correctly identified 214,837 of 215,689 non-CRC cases).

Statistical Analyses

Descriptive statistics were generated for the characteristics of the development, internal validation, and external validation cohorts. Characteristics for the internal and external validation cohorts were compared to the characteristics for the development cohort. Chi-square tests, 2-sample t-tests, and Wilcoxon rank-sum tests were used to compare proportions, mean, and median values, respectively.

Model development and validation. The model development (70%) and internal validation (30%) cohorts were created using random selection of the entire set of eligible patients undergoing surveillance colonoscopy in 2014–2019. This selection was performed using the SAS SURVEYSELECT procedure with the simple random sampling method, stratifying by the year of surveillance colonoscopy. Logistic regression was used to evaluate the univariate associations between each candidate predictor and CRC in the development data set, followed by stepwise logistic regression with an entry criterion of $P < .25$ and a stay criterion of $P < .05$ for the multivariable model. The guideline-based model (ie, patient risk status according to current postpolypectomy surveillance guidelines) and the multivariable model were applied to the internal and external validation data sets. To explore (and potentially remedy) model drift, an updated multivariable risk prediction model (*updated multivariable model*) was similarly developed using the external validation data set. All analyses used SAS software v.9.4 (SAS Institute Inc., Cary, North Carolina, US).

Model predictive performance. Model *discrimination* was assessed using the area under the receiver-operating characteristic curve [AUC]), which estimates the model's ability to discriminate between high- and low-risk CRC observations. Model *calibration* was calculated using the Hosmer–Lemeshow goodness-of-fit test; well-fitted models showing nonsignificance on the test ($P \geq .05$) indicate that modeled and observed prediction are not significantly different. The *Akaike Information Criterion* (AIC) was used to compare the performance of the prediction models.[9] Typically, if the AIC difference is larger than 2, the model with the lower AIC is considered better, and an AIC difference >10 is considered very strong evidence that the model performance is better.

Risk scoring. To facilitate the practical use of the prediction risk scores from the multivariable logistic regression model, a point system was created based on the parameter estimates for β coefficients from the model. Each predictor was standardized using the β coefficient for the covariate age, whereby each predictor was assigned a weight proportional to the value of its β coefficient in reference points; thus, each point was the equivalent of 1 year of aging. Score values were then assigned in each cohort according to the point value for each predictor and corresponding category for a patient. Risk scores from the model development cohort were divided into deciles and, using these decile score ranges as the reference, the number of surveillance colonoscopies performed, and CRCs detected were evaluated within each decile for each cohort. The process was repeated using coefficients and risk scores from the updated multivariable model performed in the external validation data set.

Results

Cohort Characteristics for Model Development and Internal Validation

A total of 52,848 patients who underwent surveillance colonoscopy in 2014–2019 were included in the model development (n = 36,994) and internal validation (n = 15,854) cohorts (Figure A1). For the cohorts combined (Table 1), the mean age was 66 ± 8 years; 41.5% were female; 16.4% were categorized as Asian or Pacific Islander, 5.6% as Black, 10.4% as Hispanic, and 50.9% as White persons. Screening was the indication for 27.7% of the index colonoscopies, and adenomas were found at 68.6% of the index procedures. The remaining surveillance colonoscopies were performed on people who had a remote history of adenoma or serrated polyp or who had a diagnosis code for history of adenoma or polyp for a colonoscopy performed outside KPNC, for which a colonoscopy or pathology report were not available. The median time interval between the index and surveillance colonoscopies was 5 years (interquartile range: 5, 6 years). Based on the prior colonoscopy findings, 72.6% of patients met the criteria for low-risk status (ie, guideline-recommended surveillance interval of ≥ 5 years), 20.4% were high risk (ie, guideline-recommended surveillance interval of <5 years), and 6.9% lacked the information needed to assign risk status. Given the development and internal validation cohorts were randomly selected from the collective pool of patients, characteristics were similar between the 2 groups. There were 114 and 43 incident CRC cases in the development and internal validation cohorts, respectively. The absolute risk of CRC was 5.43 and 5.08 per 10,000 patient-years of follow-up in the development and internal validation cohorts, respectively; patients missing the date of prior colonoscopy were excluded from the analysis.

Model Performance in the Model Development Cohort

In the model development cohort, in univariate analyses (Table 2), 7 variables were found to be significantly associated with risk of CRC at surveillance, including increasing age (per 1-year increase), ever smoked tobacco, a time

Table 2. Candidate Predictors and Their Association With CRC Diagnosed at or Within 6 mo After Surveillance Colonoscopy in the Development Cohort

Patient demographic and clinical characteristics at surveillance	Unadjusted CRC risk OR (95% CI)	Multivariable adjusted CRC risk OR (95% CI)
Age (per 1-y increase)	1.08 (1.05–1.11)	1.08 (1.05–1.11)
Sex		
Female	1.00 (reference)	
Male	1.09 (0.75–1.59)	
Race and ethnicity		
Asian or Pacific Islander	0.62 (0.34–1.12)	
Black	0.96 (0.44–2.10)	
Hispanic	0.75 (0.39–1.47)	
White	1.00 (reference)	
Other and unknown	0.89 (0.53–1.48)	
First-degree family history of CRC		
Yes	1.01 (0.63–1.62)	
No or unknown	1.00 (reference)	
Body mass index (per 1 Kg/m^2 increase)	0.99 (0.96–1.03)	
Ever smoked tobacco		
Yes	2.00 (1.39–2.89)	1.77 (1.22–2.57)
No or unknown	1.00 (reference)	1.00 (reference)
Diabetes		
Yes	1.17 (0.76–1.82)	
No	1.00 (reference)	
Index colonoscopy-related characteristics		
Performed within KPNC (and have information)		
Yes	1.00 (reference)	
No	1.15 (0.58–2.27)	
Bowel preparation		
Good or excellent	1.00 (reference)	
Poor or fair	1.37 (0.80–2.34)	
Missing	1.07 (0.63–1.84)	
Colonoscopy indication		
Screening	1.00 (reference)	
Diagnostic	0.79 (0.42–1.48)	
Positive fecal test	2.13 (1.28–3.55)	
Surveillance	1.31 (0.76–2.25)	
Unknown	1.44 (0.67–3.09)	
ADR, all indications, %		
<32.5 or missing	1.94 (1.34–2.81)	1.96 (1.35–2.84)
≥32.5	1.00 (reference)	1.00 (reference)
Time from index to surveillance colonoscopy, y		
2 to <3	1.13 (0.35–3.64)	
3 to <5	1.24 (0.78–1.96)	
5 to <7	1.00 (reference)	
7 to <10	2.08 (1.28–3.39)	
≥10	1.10 (0.27–4.53)	
Missing	1.39 (0.66–2.94)	
Index colonoscopy findings		
Any adenoma		
Yes	1.36 (0.85–2.18)	
No	1.00 (reference)	
Missing	1.45 (0.67–3.16)	
Adenoma with advanced histology		
Yes	1.31 (0.70–2.44)	
No	1.00 (reference)	
Missing	1.18 (0.59–2.33)	
Polyp size		
≥10 mm	2.01 (1.27–3.18)	2.08 (1.31–3.29)
<10 mm, no polyp, or missing	1.00 (reference)	1.00 (reference)
Number of adenomas		
0–2	1.00 (reference)	
≥3	0.88 (0.43–1.81)	
Missing	1.14 (0.57–2.25)	

Table 2. Continued		
Patient demographic and clinical characteristics at surveillance	Unadjusted CRC risk OR (95% CI)	Multivariable adjusted CRC risk OR (95% CI)
Risk status based on postpolypectomy surveillance guidelines		
High risk	1.00 (reference)	
Low risk	1.41 (0.92–2.16)	
Missing	1.25 (0.63–2.50)	
Worst finding		
Advanced adenoma (size ≥10 mm or advanced histology)	1.97 (1.12–3.46)	
Nonadvanced adenoma	1.18 (0.72–1.94)	
No adenoma	1.00 (reference)	
Missing	1.45 (0.67–3.15)	

ADR, adenoma detection rate; CI, confidence interval; CRC, colorectal cancer; KPNC, Kaiser Permanente Northern California; OR, odds ratio.

interval of 7 to ≤10 years between the index and surveillance colonoscopy, and, for the index colonoscopy, a positive fecal test indication, having the procedure performed by a physician with an all-indication ADR <32.5% or missing, having an advanced adenoma, and having a polyp size ≥10 mm. Notably, patient risk status according to current post polypectomy surveillance guidelines was not a statistically significant predictor of CRC.

In the multivariable model (Table 2), 4 variables remained statistically significant predictors of CRC at surveillance ($P < .05$), including increasing patient age (per 1-year increase), ever smoked tobacco, endoscopist all-indication ADR <32.5% or missing, and polyp size ≥10 mm.

As shown in Figure panel A, the multivariable model yielded an AUC of 0.71 (95% confidence interval [CI]: 0.67–0.76) and demonstrated good calibration by the goodness-of-fit test ($P = .49$). In comparison, the guideline-based model yielded an AUC of 0.53 (95% CI: 0.49–0.58) and demonstrated good calibration ($P = .99$). AIC values for the multivariable model and guideline-based model were 1483.4 and 1547.5, respectively, a difference of 64.1.

Using the multivariable model, 29 of 114 patients with CRC (25.4%) were found in the top decile and 68 of 114 patients with CRC (59.6%) in the top 3 deciles of risk score (Table 3). With 3542 and 10,929 surveillance colonoscopies performed in the top decile and top 3 deciles, respectively, this equates to 8.2 and 6.2 CRCs detected per 1000 surveillance colonoscopies performed.

Model Performance in the Internal Validation Cohort

Applying the multivariable model to the internal validation cohort of 15,854 patients yielded an AUC of 0.73 (95% CI: 0.66–0.81) and demonstrated good calibration ($P = .79$) (Figure panel B). In comparison, the guideline-based model yielded an AUC of 0.52 (95% CI: 0.45–0.60) and demonstrated good calibration ($P = .18$). AIC values for the multivariable model and guideline-based model were 566.9 and 597.6, respectively, a difference of 30.7.
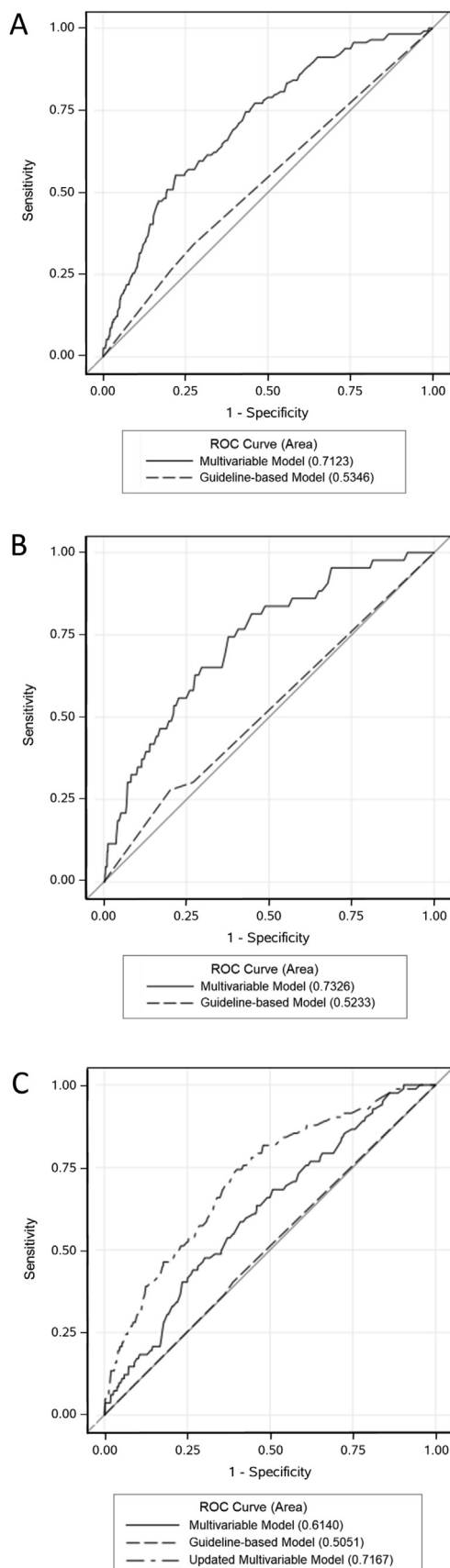
Using the multivariable model, 16 of 43 CRC cases (37.2%) were found in the top decile and 30 of the 43 CRC cases (70%) in the top 3 deciles of risk score, respectively (Table 3). With 1454 and 4568 surveillance colonoscopies performed in the top decile and top 3 deciles, respectively, this equates to 11.0 and 6.6 CRCs detected per 1000 surveillance colonoscopies performed.

Model Performance in the External Validation Cohort

In the external validation cohort, 38,242 patients underwent a surveillance colonoscopy, and 82 individuals were diagnosed with CRC at or within 6 months after the procedure (Figure A2). The absolute risk of CRC in the external validation cohort was 3.90 per 10,000 patient-years of follow-up and patients missing the date of prior colonoscopy were excluded from the analysis. Differences in the characteristics of the external validation cohort compared to the development cohort included greater proportions of patients with index colonoscopy endoscopist ADR ≥32.5% (the median all-indication ADR was 46%); adequate bowel preparation; complete examination; <5 years as the interval between the index and surveillance procedure; and having 3 or more polyps at the index colonoscopy (Table 1). In contrast, a lower proportion had a polyp ≥10 mm in size or advanced histology at the index colonoscopy. The absolute risk of CRC at surveillance colonoscopy was also lower in the external validation cohort.

Applying the multivariable model to the external validation data set yielded an AUC of 0.61 (95% CI: 0.56–0.67) and demonstrated good calibration ($P = .87$) (Figure panel C). In comparison, the guideline-based model yielded an AUC of 0.51 (95% CI: 0.45–0.56) and demonstrated good calibration ($P = .53$). AIC values for the multivariable model and guideline-based model were 1160.9 and 1175.6, respectively, a difference of 14.7.

Using the multivariable model, 10 of 82 CRC cases (12.2%) were found in the top decile and 33 of the 82 CRC cases (40.2%) in the top 3 deciles of risk score (Table 3).

A

ROC Curve (Area)
Multivariable Model (0.7123)
Guideline-based Model (0.5346)

B

ROC Curve (Area)
Multivariable Model (0.7326)
Guideline-based Model (0.5233)

C

ROC Curve (Area)
Multivariable Model (0.6140)
Guideline-based Model (0.5051)
Updated Multivariable Model (0.7167)

With 2525 and 9100 surveillance colonoscopies performed in the top decile and top 3 deciles, respectively, this equates to 4.0 and 3.6 CRCs detected per 1000 surveillance colonoscopies.

Given the observed performance drift of the multivariable model in the external validation cohort, the multivariable model was updated. In univariate analyses (Table 4), 10 variables were found to be significantly associated with risk of CRC at surveillance, including increasing age (per 1-year increase), unknown race or ethnicity, a time interval of 7 to $\leq$10 or >10 years between the index and surveillance colonoscopies, and, for the index colonoscopy, unknown bowel preparation, a positive fecal test or surveillance indication, having the procedure performed by a physician with an all-indication ADR <32.5% or missing as well as ADR <37.5% or missing, having an advanced adenoma, and having an adenoma with advanced histology. Again, notably, guideline-based risk status was not a statistically significant predictor of CRC.

In the updated multivariable model (Table 4), 4 variables remained statistically significant predictors of CRC at surveillance ($P < .05$), including increasing patient age (per 1-year increase), and for the index colonoscopy, a positive fecal test indication, endoscopist all-indication ADR <37.5% or missing, and adenoma with advanced histology.

As shown in Figure panel C, in the external validation data set, the updated multivariable model yielded an AUC of 0.72 (95% CI: 0.66–0.77) and demonstrated good calibration ($P = .42$).

Discussion

There are circumstances when surveillance colonoscopy demand exceeds capacity; this was the case during and following the COVID-19 pandemic.[3–5] In such circumstances, prediction modeling may be a helpful tool to ensure patients awaiting surveillance colonoscopy and at high risk for CRC are prioritized for the procedure. In the development and internal validation cohorts of over 52,000 patients who received a surveillance colonoscopy, the multivariable model consisting of polyp size $\geq$10 mm, endoscopist ADR <32.5% or missing, increasing patient age, and ever smoked tobacco demonstrated good discrimination with AUCs in the

Figure. Model performance in the model development (panel A), internal validation (panel B), and external validation (panel C) cohorts based on area under the receiver-operating characteristic (ROC) curves. The *guideline-based model* is based on the guideline-recommended surveillance interval dichotomized as low- or high-risk (ie, recommended interval $\geq$5 vs <5 years). The *multivariable model* includes age (in 1-year increments), index endoscopist all-indication ADR <32.5% or missing, index colonoscopy polyp size $\geq$10 mm, and ever smoked. The *updated multivariable model* includes patient age (in 1-year increments), a positive fecal test indication on the index colonoscopy, index endoscopist all-indication ADR <37.5% or missing, and adenoma with advanced histology.

Table 3. Distribution of Surveillance Colonoscopies and CRCs Detected at or Within 6 mo After Surveillance Colonoscopy, Stratified by Risk Score Decile

CRC risk score decile	Model development cohort: multivariable model		Internal validation cohort: multivariable model		External validation cohort: multivariable model		External validation cohort: updated multivariable model	
	Colo n	CRC n (%)[a]	Colo n	CRC n (%)[a]	Colo n	CRC n (%)[a]	Colo n	CRC n (%)[a]
First	3571	2 (1.8)	1502	2 (4.7)	4437	2 (2.4)	3900	1 (1.2)
Second	3891	3 (2.6)	1610	1 (2.3)	4874	9 (11.0)	3950	5 (6.1)
Third	3534	5 (4.4)	1454	0 (0.0)	3775	6 (7.3)	3885	2 (2.4)
Fourth	3842	6 (5.3)	1688	3 (7.0)	4144	8 (9.8)	3349	3 (3.7)
Fifth	3620	8 (7.0)	1606	1 (2.2)	3724	7 (8.5)	3827	4 (4.9)
Sixth	3740	11 (9.6)	1674	5 (11.6)	4186	10 (12.2)	4376	7 (8.5)
Seventh	3867	11 (9.6)	1752	4 (9.3)	4002	7 (8.5)	3794	13 (15.9)
Eighth	3398	10 (8.8)	1470	5 (11.6)	3327	16 (19.5)	3606	9 (11.0)
Ninth	3989	29 (25.0)	1644	6 (14.0)	3248	7 (8.5)	3870	14 (17.1)
10th	3542	29 (25.0)	1454	16 (37.2)	2525	10 (12.2)	3685	24 (29.3)
Total	36,994	114 (100)	15,854	43 (100)	38,242	82 (100)	38,242	82 (100)

Colo, colonoscopy; CRC, colorectal cancer.
[a]Reflects the percent of total CRC, cases in the cohort accounted for in the CRC, risk score decile.

range of 0.71–0.73 and good calibration, and approximately 57%–70% of detected CRC cases were found in the top 3 deciles of risk score. In contrast, the guideline-based model which categorized patients as either high or low risk based on polyp finding at the index colonoscopy offered little predictive value for patients awaiting surveillance, with AUCs in the range of 0.52–0.53. Performance of the multivariable model drifted when applied to the external validation cohort from a later period but improved with model updating.

The current study is unique in seeking to develop a model to prioritize those most likely to be diagnosed with CRC among patients awaiting surveillance colonoscopy. However, the predictive factors identified and model performance are generally consistent with prior studies that sought to improve upon guideline-based post polypectomy risk stratification for advanced adenoma or advanced neoplasia at surveillance colonoscopy.[10–14] While the multivariable model in the current study was designed exclusively for the more definitive outcome of CRC, the development and internal validation AUCs of 0.71 and 0.73 were similar to the range of AUCs reported for these other models (0.62–0.71).[10–14] Also, the independent predictors of CRC identified in the multivariable model in the current study have also been reported in other models for either advanced adenoma or advanced neoplasia prediction, specifically for increasing patient age,[10,11,13,14] current or former smoker,[10] adenoma size ≥10 mm,[10,12,14] and endoscopist all-indication ADR.[10]

The finding that lower endoscopist all-indication ADR was an independent predictor of CRC at surveillance colonoscopy is consistent with our prior work demonstrating that ADR is strongly inversely associated with risk of postcolonoscopy CRC and related death.[15,16] The prediction model findings in the current study, along with those of Gupta et al for the outcome of metachronous advanced neoplasia,[10] provide further support for efforts to improve endoscopist ADRs,[17] and the use of all-indication ADRs as a simpler and easier-to-calculate alternative to screening ADR,[18,19] including in risk prediction models. In our setting, endoscopist ADRs have improved over the last decade,[20] so future versions of the model are likely to need additional updates until ADR reaches a plateau, beyond which further improvements in ADR do not decrease risk of postcolonoscopy CRC.

A better understanding of CRC risk among patients awaiting colonoscopy surveillance through prediction modeling also has the potential to improve selection of patients who may be appropriate for noninvasive methods of postpolypectomy surveillance. For example, the noninvasive fecal immunochemical test is the recommended approach for patients with 1-2 low-risk adenomas in the Canadian province of Ontario,[21] and is the subject of an ongoing clinical trial (ClinicalTrials.gov ID NCT05612347).

A key strength of the current study is completion of both an internal and an external validation and updating the multivariable model. In this case, the external validation of

Table 4. Candidate Predictors and Their Association With CRC Diagnosed at or Within 6 mo After Surveillance Colonoscopy in the External Validation Cohort

Patient demographic and clinical characteristics at surveillance	Unadjusted CRC risk OR (95% CI)	Adjusted CRC risk: *updated multivariable model* OR (95% CI)
Age (per 1-y increase)	1.05 (1.02–1.08)	1.06 (1.03–1.09)
Sex		
Female	1.00 (reference)	
Male	1.04 (0.67–1.62)	
Race and ethnicity		
Asian or Pacific Islander	0.69 (0.38–1.25)	
Black	0.90 (0.38–2.10)	
Hispanic	0.57 (0.26–1.27)	
White	1.00 (reference)	
Other and unknown	0.37 (0.19–0.73)	
First-degree family history of CRC		
Yes	1.01 (0.58–1.77)	
No or unknown	1.00 (reference)	
Body mass index (per 1 Kg/m^2 increase)	1.02 (0.98–1.05)	
Ever smoked tobacco		
Yes	1.14 (0.72–1.81)	
No or unknown	1.00 (reference)	
Diabetes		
Yes	1.41 (0.87–2.28)	
No	1.00 (reference)	
Index colonoscopy-related characteristics		
Performed within KPNC (and have information)		
Yes	1.00 (reference)	
No	1.47 (0.46–4.65)	
Bowel preparation		
Good or excellent	1.00 (reference)	
Poor or fair	1.03 (0.45–2.38)	
Missing	2.27 (1.13–4.57)	
Colonoscopy indication		
Screening	1.00 (reference)	
Diagnostic	1.60 (0.66–3.86)	
Positive fecal test	4.72 (2.26–9.87)	2.71 (1.71–4.28)
Surveillance	2.28 (1.07–4.87)	
Unknown	3.38 (0.91–12.49)	
All indications except positive fecal test		1.00 (reference)
ADR, all indications, %		
<32.5 or missing	1.83 (1.14–2.95)	———
≥32.5	1.00 (reference)	———
ADR, all indications, %		
<37.5 or missing	———	2.68 (1.73–4.14)
≥37.5	———	1.00 (reference)
Time from index to surveillance colonoscopy, years		
2 to <3	1.88 (0.64–5.57)	
3 to <5	1.52 (0.85–2.72)	
5 to <7	1.00 (reference)	
7 to <10	2.69 (1.39–5.23)	
≥10	4.72 (2.12–10.53)	
Missing	2.37 (0.70–8.07)	
Index colonoscopy findings		
Any adenoma		
Yes	0.94 (0.56–1.59)	
No	1.00 (reference)	
Missing	1.40 (0.41–4.75)	
Adenoma with advanced histology		
Yes	2.40 (1.30–4.45)	2.16 (1.15–4.09)
No	1.00 (reference)	———
Missing	1.61 (0.51–5.12)	———
No or missing		1.00 (reference)
Polyp size		
≥10 mm	1.17 (0.60–2.27)	
<10 mm, no polyp, or missing	1.00 (reference)	

Table 4. Continued

Patient demographic and clinical characteristics at surveillance	Unadjusted CRC risk OR (95% CI)	Adjusted CRC risk: *updated multivariable model* OR (95% CI)
Number of adenomas		
0–2	1.00 (reference)	
≥3	0.99 (0.61–1.61)	
Missing	1.46 (0.46–4.67)	
Risk status based on post polypectomy surveillance guidelines		
High risk	1.03 (0.66–1.63)	
Low risk or none	1.00 (reference)	
Missing	1.48 (0.46–4.77)	
Worst finding		
Advanced adenoma (size ≥10 mm or advanced histology)	1.34 (0.69–2.61)	
Nonadvanced adenoma	0.84 (0.49–1.46)	
No adenoma	1.00 (reference)	
Missing	1.40 (0.41–4.75)	

ADR, adenoma detection rate; CI, confidence interval; CRC, colorectal cancer; KPNC, Kaiser Permanente Northern California; OR, odds ratio.

prediction models was performed using a data set from the same health system, but at a later time period (ie, temporal validation), consistent with Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis guideline recommendations.[22] A recognized limitation of the risk prediction modeling literature is that external validation of models occurs infrequently, and when external validation is performed, discrimination is often found to be lower.[23] In the current study, model performance was lower when the multivariable model was applied to an external validation cohort of about 38,000 patients who received a surveillance colonoscopy during the early years of the COVID-19 pandemic. This finding underscores that prediction modeling studies that only report AUCs based on internal validation testing may be overestimating model performance in clinical practice. The lower performance at external validation is likely due to the well-documented phenomenon of model drift, where model performance declines in different populations or time periods.[24–29] For example, the temporal improvements in ADR among providers for patients in the external validation cohort compared to the model development cohort may have altered the threshold at which this variable was most predictive of CRC and therefore weakened the predictive power for this variable at the level specified in the model. Further, relationships between the variables may have changed significantly during the COVID-19 pandemic, as the pandemic modified who was seeking or referred for care.[30] Also, the absolute risk of CRC was lower in the external validation cohort compared to the model development cohort, likely due in part to improvements in endoscopist ADRs. These findings highlight the need for continual updating and validation of models and risk scores to ensure their utility for clinicians.[31,32] Optimal updating methods are an active area of research.[33] In the present study, the updated multivariable model followed the same methods as that for the training data, and recovered the performance of the earlier (multivariable) model, with a high AUC and good calibration. Stepwise logistic regression is a well-documented procedure that can be implemented in free software, making the task of updating the model on a regular—perhaps annual—basis reasonably straightforward.

Other strengths of the study include prediction model development and internal validation using a large sample of surveillance colonoscopies from a population-based integrated health-care setting, evaluating a wide range of candidate predictors including all-indication ADR, use of validated methods for capturing adenoma findings, focusing specifically on CRC as the study outcome with the intent to prioritize patients awaiting surveillance colonoscopy, and ascertaining CRC detection using a validated cancer registry or manual chart review. A further strength is that cohorts retained members who met the eligibility criteria but whose index colonoscopy was not performed within KPNC and who were therefore missing information for predictors related to index colonoscopy; this reflects a common real-world situation where patients are missing data on potential predictors of colonoscopy outcome.

Several limitations must also be considered. First, this was an observational study and not a randomized trial. Therefore, findings are subject to the limitations of observational study designs, such as confounding by variables unaccounted for and possible misclassification of predictors. Second, several factors could limit the generalizability of the study findings including the integrated health-care setting, the fact that the study population is medically insured and does not include extremes of poverty and wealth, the health system has lower rates of smoking than the surrounding population due to active promotion of smoking cessation programs, and the diversity and granularity of data in this

setting may be greater than for other settings. Third, the development and internal validation data sets were comprised of a random split of a larger data set which reduced the sample size for model development. Fourth, model results may have been impacted by the level of detail for predictor variables as recorded in electronic medical records, and potentially important candidate predictors may have been missed. Fifth, individual polyps excised at colonoscopy are often placed into separate pathology jars for histological examination, but multiple polyps taken from the same location may be placed in the same jar. Thus, the number of pathology jars with a confirmed adenoma is a surrogate for the total number of adenomas detected at colonoscopy, but the count is subject to under-reporting. The exact number of adenomas in each jar is subject to misclassification due to limitations in our data, which may explain why adenoma number was not a significant predictor in our models. Sixth, some predictors had missing data, although a *missing* category was included to account for this; nonetheless, misclassification may have influenced the results. Seventh, as noted above, the time interval for the external validation included the COVID-19 pandemic, when there were fewer colonoscopies performed and higher-risk patients, such as those with symptoms, a positive noninvasive test, or high-risk surveillance, were prioritized.[30] Supporting this possibility, about 47% of the external validation cohort had a time interval between the index and surveillance procedures of <5 years, compared with 28% of those in the model development and internal validation cohorts. Also, colonoscopy quality improvement efforts have led to increased endoscopist ADRs,[16] and consequently, the proportion of index colonoscopies that detected 3 or more adenomas was higher in the external validation cohort than the other 2 cohorts (28% vs 8%), possibly reflecting improved detection without a true change in risk. Eighth, while patients with a distant history of adenomas or serrated polyps were included in the study, the study only considered variables associated with the index colonoscopy immediately preceding the surveillance colonoscopy. Some patients may have had a colonoscopy in the more distant past with more advanced findings than seen on the most recent index colonoscopy. In addition, data were unavailable on the total number of prior colonoscopies for each patient as procedures could have been performed outside the health system; thus, we were unable to evaluate the number of prior colonoscopies as a candidate predictor. Ninth, in considering polyp size as a candidate predictor, we investigated the frequency of polyps ≥ 10 mm vs >20 mm in size and found there were too few of the latter to include it as a candidate predictor in the models. Finally, the appropriate next step is to repeat the validation step in another external cohort.

Conclusion

The clinical implication of the present study is that at those times when surveillance colonoscopy demand exceeds capacity and patients are on a waiting list for the procedure, risk prediction modeling may be useful for prioritizing patients at highest CRC risk using variables that are routinely collected in clinical practice or for quality reporting. These patients could be considered for more intensive efforts to ensure prompt surveillance colonoscopy.

Supplementary Materials

Material associated with this article can be found in the online version at https://doi.org/10.1016/j.gastha.2024.03.008.

References

1. US Preventive Services Task Force, Bibbins-Domingo K, Grossman DC, et al. Screening for colorectal cancer: US preventive Services task Force recommendation Statement. JAMA 2016;315:2564–2575.
2. Gupta S, Lieberman D, Anderson JC, et al. Recommendations for follow-up after colonoscopy and polypectomy: a consensus update by the US multi-Society Task Force on Colorectal Cancer. Gastroenterology 2020;158:1131–1153.e5.
3. Tinmouth J, Dong S, Stogios C, et al. Estimating the backlog of colonoscopy due to coronavirus disease 2019 and comparing strategies to recover in Ontario, Canada. Gastroenterology 2021;160:1400–1402.e1.
4. Lantinga MA, Theunissen F, Ter Borg PCJ, et al. Impact of the COVID-19 pandemic on gastrointestinal endoscopy in The Netherlands: analysis of a prospective endoscopy database. Endoscopy 2021;53:166–170.
5. Rutter MD, Brookes M, Lee TJ, et al. Impact of the COVID-19 pandemic on UK endoscopic activity and cancer detection: a National Endoscopy Database Analysis. Gut 2021;70:537–543.
6. Gordon NP. How does the adult kaiser permanente membership in northern California compare with the larger community? Oakland, CA: Kaiser Permanente Division of Research, 2006.
7. Davis AC, Voelkel JL, Remmers CL, et al. Comparing kaiser permanente members to the general population: implications for generalizability of research. Perm J 2023;27:87–98.
8. Lee JK, Jensen CD, Lee A, et al. Development and validation of an algorithm for classifying colonoscopy indication. Gastrointest Endosc 2015;81:575–582.e4.
9. Burnham KP, Anderson DR. Multimodel inference: understanding AIC and BIC in model selection. Sociol Methods Res 2004;33:261–304.
10. Gupta S, Earles A, Bustamante R, et al. Adenoma detection rate and clinical characteristics influence advanced neoplasia risk after colorectal polypectomy. Clin Gastroenterol Hepatol 2023;21:1924–1936.e9.
11. Laiyemo AO, Murphy G, Albert PS, et al. Postpolypectomy colonoscopy surveillance guidelines: predictive accuracy for advanced adenoma at 4 years. Ann Intern Med 2008;148:419–426.
12. van Heijningen EM, Lansdorp-Vogelaar I, Kuipers EJ, et al. Features of adenoma and colonoscopy associated

with recurrent colorectal neoplasia based on a large community-based study. Gastroenterology 2013; 144:1410–1418.

13. Lee JY, Park HW, Kim MJ, et al. Prediction of the risk of a metachronous advanced colorectal neoplasm using a novel scoring system. Dig Dis Sci 2016;61: 3016–3025.

14. Liu L, Messer K, Baron JA, et al. A prognostic model for advanced colorectal neoplasia recurrence. Cancer Causes Control 2016;27:1175–1185.

15. Corley DA, Jensen CD, Marks AR, et al. Adenoma detection rate and risk of colorectal cancer and death. N Engl J Med 2014;370:1298–1306.

16. Schottinger JE, Jensen CD, Ghai NR, et al. Association of physician adenoma detection rates with postcolonoscopy colorectal cancer. JAMA 2022;327:2114–2122.

17. Shaukat A, Tuskey A, Rao VL, et al. Interventions to improve adenoma detection rates for colonoscopy. Gastrointest Endosc 2022;96:171–183.

18. Kaltenbach T, Gawron A, Meyer CS, et al. Adenoma detection rate (ADR) irrespective of indication is comparable to screening ADR: implications for quality monitoring. Clin Gastroenterol Hepatol 2021; 19:1883–1889.e1.

19. Corley DA, Jensen CD, Chubak J, et al. Evaluating different approaches for calculating adenoma detection rate: is screening colonoscopy the gold standard? Gastroenterology 2023;165:784–787.e4.

20. Corley DA, Jensen CD, Lee JK, et al. Impact of a scalable training program on the quality of colonoscopy performance and risk of post-colonoscopy colorectal cancer. Gastrointest Endosc 2023;98:609–617.

21. Dube C. ColonCancerCheck recommendations for post-polypectomy surveillance. Cancer Care Ontario. https://www.cancercareontario.ca/en/guidelines-advice/types-of-cancer/38506. Accessed October 26, 2023.

22. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1–W73.

23. Siontis GC, Tzoulaki I, Castaldi PJ, et al. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. J Clin Epidemiol 2015;68:25–34.

24. Toll DB, Janssen KJ, Vergouwe Y, et al. Validation, updating and impact of clinical prediction rules: a review. J Clin Epidemiol 2008;61:1085–1094.

25. Davis SE, Lasko TA, Chen G, et al. Calibration drift in regression and machine learning models for acute kidney injury. J Am Med Inform Assoc 2017;24:1052–1061.

26. Davis SE, Greevy RA Jr, Lasko TA, et al. Detection of calibration drift in clinical prediction models to inform model updating. J Biomed Inform 2020;112:103611.

27. Hickey GL, Grant SW, Murphy GJ, et al. Dynamic trends in cardiac surgery: why the logistic EuroSCORE is no longer suitable for contemporary cardiac surgery and implications for future risk models. Eur J Cardiothorac Surg 2013;43:1146–1152.

28. Davis SE, Lasko TA, Chen G, et al. Calibration drift among regression and machine learning models for hospital mortality. AMIA Annu Symp Proc 2017; 2017:625–634.

29. Moons KG, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. Heart 2012;98:691–698.

30. Lee JK, Lam AY, Jensen CD, et al. Impact of the COVID-19 pandemic on fecal immunochemical testing, colonoscopy services, and colorectal neoplasia detection in a large United States community-based population. Gastroenterology 2022;163:723–731.e6.

31. Davis SE, Brown JR, Dorn C, et al. Maintaining a national acute kidney injury risk prediction model to support local quality benchmarking. Circ Cardiovasc Qual Outcomes 2022;15:e008635.

32. Minne L, Eslami S, de Keizer N, et al. Effect of changes over time in the performance of a customized SAPS-II model on the quality of care assessment. Intensive Care Med 2012;38:40–46.

33. Lee CS, Lee AY. Clinical applications of continual learning machine learning. Lancet Digit Health 2020; 2:e279–e281.

Authors' Contributions:
Theodore R. Levin, MD (Conceptualization: lead; Funding Acquisition: lead; Formal Analysis: supporting; Writing – Original Draft Preparation: equal; Writing – Review & Editing: equal). Christopher D. Jensen, PhD, MPH (Conceptualization: equal; Formal Analysis: supporting; Writing – Original Draft Preparation: lead; Writing – Review & Editing: lead). Amy R. Marks, MPH (Conceptualization: equal; Curation: lead; Formal Analysis: lead; Writing – Original Draft Preparation: supporting; Writing – Review & Editing: supporting). David Schlessinger, MS (Methodology: supporting; Writing – Review & Editing: supporting). Vincent Liu, MD (Methodology: supporting; Writing – Review & Editing: supporting). Natalia Udaltsova, PhD (Resources: supporting). Jessica Badalov, MPH (Project Administration: lead; Writing – Review & Editing: supporting). Evan Layefsky, BS (Project Administration: supporting; Writing – Review & Editing: supporting). Douglas A. Corley, MD, PhD (Conceptualization: supporting; Funding Acquisition: supporting; Writing – Review & Editing: supporting). Joshua R. Nugent, PhD (Writing – Original Draft Preparation: supporting; Writing – Review & Editing: supporting). Jeffrey K. Lee, MD, MPH (Conceptualization: equal; Funding Acquisition: supporting; Writing – Review & Editing: supporting).