

Measuring the observer (Hawthorne) effect on adenoma detection rates



Authors

Mahsa Taghiakbari¹, Diana Elena Coman², Mark Takla³, Alan Barkun¹, Mickael Bouin¹, Simon Bouchard^{4,1}, Eric Deslandres^{4,1}, Sacha Sidani^{4,1}, Daniel von Renteln¹

Institutions

- 1 Gastroenterology, Centre Hospitalier de l'Université de Montréal (CHUM), Montreal, Canada
- 2 Internal Medicine, Centre Hospitalier de l'Université de Montréal (CHUM), Montreal, Canada
- 3 Faculty of Medicine, University of Montreal Hospital Centre, Montreal, Canada
- 4 Gastroenterology, Centre de Recherche de l'Université de Montréal (CHUM), Montreal, Canada

Key words

Polyps / adenomas / ..., Endoscopy Lower GI Tract, Colorectal cancer, CRC screening, Endoscopic resection (polypectomy, ESD, EMRc, ...)

received 5.11.2022

accepted after revision 13.7.2023

accepted manuscript online 17.7.2023

Bibliography

Endosc Int Open 2023; 11: E908–E919

DOI 10.1055/a-2131-4797

ISSN 2364-3722

© 2023. The Author(s).

This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Georg Thieme Verlag KG, Rüdigerstraße 14,
70469 Stuttgart, Germany

Additional material is available at
<https://doi.org/10.1055/a-2131-4797>

Corresponding author

Dr. Mahsa Taghiakbari, Centre Hospitalier de l'Université de Montréal (CHUM), Gastroenterology, Montreal, Canada
mahtakbar@gmail.com
mahsa.taghiakbari@gmail.com

ABSTRACT

Background and study aims An independent observer can improve procedural quality. We evaluated the impact of the observer (Hawthorne effect) on important quality metrics during colonoscopies.

Patients and Methods In a single-center comparative study, consecutive patients undergoing routine screening or diagnostic colonoscopy were prospectively enrolled. In the index group, all procedural steps and quality metrics were observed and documented, and the procedure was video recorded by an independent research assistant. In the reference group, colonoscopies were performed without independent observation. Colonoscopy quality metrics such as polyp, adenoma, serrated lesions, and advanced adenoma detection rates (PDR, ADR, SLDR, AADR) were compared. The probabilities of increased quality metrics were evaluated through regression analyses weighted by the inversed probability of observation during the procedure.

Results We included 327 index individuals and 360 referents in the final analyses. The index group had significantly higher PDRs (62.4% vs. 53.1%, $P=0.02$) and ADRs (39.4% vs. 28.3%, $P=0.002$) compared with the reference group. The SLDR and AADR were not significantly increased. After adjusting for potential confounders, the ADR and SLDR were 50% (relative risk [RR] 1.51; 95% CI 1.05–2.17) and more than twofold (RR 2.17; 95% CI 1.05–4.47) more likely to be higher in the index group than in the reference group.

Conclusions The presence of an independent observer documenting colonoscopy quality metrics and video recording the colonoscopy resulted in a significant increase in ADR and other quality metrics. The Hawthorne effect should be considered an alternative strategy to advanced devices to improve colonoscopy quality in practice.

Introduction

The effectiveness of colonoscopy screening in preventing colorectal cancer (CRC) is directly linked to its procedural quality [1]. The adenoma detection rate (ADR), defined as the proportion of colonoscopy procedures with at least one adenoma detected, serves as the paramount operator-dependent quality metric [2]. The American College of Gastroenterology (ACG) and the American Society for Gastrointestinal Endoscopy (ASGE) recommend ADR benchmarks of 25% for all patients and sex-specific rates of 30% for men and 20% for women [2]. A higher ADR is associated with a lower risk of interval cancer and reduced mortality [3]. Other known quality metrics that improve ADR are related to the endoscopist's level of experience and procedural quality, including withdrawal time, cecal intubation rate, and bowel preparation quality [2,4]. Recent advances in endoscopy technology, such as high-definition imaging or computer-assisted colonoscopy with artificial intelligence (AI) support, have shown promising results for improving ADR [5]. Given the variable ADR among endoscopists and its potential to encourage a "one-and-done" approach, it cannot sufficiently reflect endoscopists' performance. Therefore, other quality metrics such as advanced adenoma detection rate (AADR) were proposed to mitigate the limitations of ADR [4, 6]. The value of these markers is, however, unknown, and they are not recommended by the colonoscopy guidelines as quality markers with an established threshold.

The presence of an independent observer (e.g. nurses, fellows, or technicians) during a medical procedure, known as the Hawthorne effect, can influence procedural quality. This effect is known to improve health-related outcomes in different disciplines. For example, a Canadian study showed an approximately threefold increase in hand hygiene dispenser use when an observer was present compared to when no observer was in sight [7]. The Hawthorne effect could also be an important factor influencing colonoscopy quality, including ADR, but available results and studies are limited or controversial [4,6]. Therefore, we compared colonoscopy quality metrics in patients undergoing outpatient colonoscopies under stringent observer conditions, involving observing and documenting all procedure steps and video-recording each colonoscopy in full length by an independent research assistant. Findings were compared with a reference group of patients who underwent routine colonoscopies without any procedural observation or documentation. We hypothesized that colonoscopies performed in the observer (index) group would have higher ADR and other quality metrics than routine colonoscopies in the reference group, conducted by the same endoscopists at the same institution for similar indications.

Patients and methods

Study design

We conducted a comparative study of patients aged 45 to 80 years who underwent an outpatient colonoscopy at the Centre Hospitalier de l'Université de Montréal (CHUM). The "observed" index cohort was prospectively collected consecutive individ-

uals within a project focused on obtaining full-length annotated video recordings of colonoscopy procedures for the development of an AI system [8]. The colonoscopy procedures were recorded, and a research assistant documented all procedure steps in real time, including the detection of lesions, withdrawal time, and identification of anatomical landmarks. However, no additional modalities, such as computer-assisted detection, or image-enhancing techniques, were employed to enhance the chances of polyp or adenoma detection. Endoscopists conducted the colonoscopies while being aware that all procedure quality metrics and characteristics would be documented by a research assistant, and the recorded videos were later utilized for analysis of procedural quality.

Written informed consent was obtained from all patients in the index group. The Institutional Review Board approval for this study protocol was granted, including waiving the need for informed consent for the retrospective reference group (CRCHUM IRB#21.045).

Study participants

The index group was selected from a cohort of patients who underwent screening, surveillance, or diagnostic colonoscopy procedures performed by four attending endoscopists between January and November 2021 at CHUM. Referents were selected by reviewing the electronic medical records of patients who underwent an outpatient colonoscopy by the same attending endoscopists at the same center between January and December 2020. Only the colonoscopies performed without the presence of a research assistant and any form of real-time video recording were included. All eligible colonoscopies, regardless of the indication or patient family history of CRC, were included. The exclusion criteria are shown in the list below. There was no overlap between the index and reference groups of patients.

Exclusion criteria

Indications for colonoscopy with high likelihood of polyp or neoplastic lesion discovery (e.g., referral for polypectomy procedures, previous computed tomography colonography, or other imaging indicating a polypoid colorectal lesion):

- Incomplete colonoscopies
- Known inflammatory bowel disease
- Active colitis
- Coagulopathy
- Familial polyposis syndrome
- Poor general health (American Society of Anesthesiologists [ASA] physical status class >III, including ASA class IV & V)

Colonoscopy procedure

Patients in both the index and reference groups received standard bowel preparation regimens and sedation [9]. All colonoscopies were performed using the same platform of high-definition video endoscopy (Olympus colonoscopes 190 series; Olympus Corp., Center Valley, Pennsylvania, United States). All detected polyps were resected using standard instruments and techniques, based on endoscopist discretion. Following polypectomy, all resected polyp specimens underwent histopathological evaluation by an institutional pathologist according to the

current institutional standards [10]. Polyps with tubulovillous or villous histology, traditional serrated adenomas, and any polyp histology showing high-grade dysplasia or cancer were classified as advanced pathology [11].

In the index group, video recordings were started prior to the insertion of the colonoscope into the patient's rectum. A stopwatch function was initiated by the research assistant upon insertion to accurately document the withdrawal time and moments of landmark detections (e.g., appendiceal orifice, ileocecal valve, polyps).

Data collection

► **Table 1** displays the collected data, including patient, procedure, and polyp/polypectomy characteristics that were assessed in the study [12]. Patient and procedure characteristics were extracted from the study case report forms for the index group and from electronic patient records for the reference group.

Study outcomes

The primary outcome was the ADR in both the index and reference groups, defined as the proportion of colonoscopies with at least one adenoma detected. Secondary outcomes included: polype detection rate, defined as the proportion of colonoscopies with at least one detected polyp of any pathology type; serrated lesion detection rate (SLDR), defined as the proportion of colonoscopies with at least one serrated lesion detected; AADR, defined as the proportion of colonoscopies with at least one advanced adenoma detected. Furthermore, we evaluated the rates of MYH-associated polyposis and mean number of adenomatous and sessile serrated polyps detected per patient (MASP), and the withdrawal time in the index and reference groups. Finally, we conducted a sensitivity analysis to compare the quality metrics between index patients and referents with positive and negative fecal immunochemical test (FIT) results.

In addition, we estimated the probability of an increase in detection rates and mean detected adenoma (and serrated lesions) per patient. We also evaluated the random effect of the endoscopists on these quality metrics, considering the fixed effect of the group.

Statistical analyses

Descriptive statistics were calculated for categorical and continuous variables as frequencies, and median (interquartile range) or mean (SD), respectively. Baseline characteristics and detection rates were compared by chi-squared or two-tailed Fisher's exact tests for the categorical variables, and Mann-Whitney U and the Kruskal-Wallis tests for continuous variables, as appropriate.

To control for patient- and procedure-related confounding factors, a propensity score was estimated to balance baseline patient and procedural characteristics between index and reference groups using logistic regression, with the presence of a research assistant as the dependent variable and the following variables as the independent variables: sex (male vs. female), age (continuous), colonoscopy indications (screening, adenoma or CRC surveillance, positive FIT, diarrhea, anemia or

► **Table 1** Data collected in the index and reference groups.

Category	Characteristics
Patient characteristics	Age
	Sex
	Procedure date (month)
	Procedure time (morning/afternoon)
	Endoscopist name
	American Society of Anesthesiologists [ASA] classification
Procedure characteristics	Colonoscopy indication
	Sedation
	Boston bowel preparation scale score (adequate or inadequate)
	Cecal intubation (surrogate for complete colonoscopy)
Polyp and polypectomy characteristics	Total withdrawal time (precisely calculated in index group, self-reported in reference group)
	Number of identified polyps
	Location (ascending, hepatic flexure, transverse, splenic flexure, descending, sigmoid, rectum)
	Size
	Morphology (polypoid/non-polypoid according to Paris classification [9])

bleeding, other [such as a change in bowel habit]), first-degree family history of CRC, endoscopist (considering the ADRs among endoscopists with various level of experience), period of colonoscopy (performed in the same month), procedure time (morning vs. afternoon procedure, considering lower afternoon ADR due to endoscopist fatigue), and physician-assessed bowel preparation quality (adequate vs. inadequate). Then, the inverse of propensity score weighting (IPTW) was used to weigh each individual. In these analyses, we included covariates mentioned above because they are known as risk factors for the polyp detection rate (PDR) and ADR [13, 14, 15, 16, 17, 18, 19, 20, 21, 22]. Previous studies have consistently shown that the ADR is higher in surveillance colonoscopies than screening colonoscopies, and screening ADR is higher than diagnostic ADR [23, 24]. The current recommendation of the ASGE/ACG is limited to the colonoscopy-naïve population with no or negative FIT or fecal occult blood test (FOBT). The target ADR in colonoscopies with the diagnostic indication (i. e., positive FIT) has been meager, although the estimated ADR in diagnostic colonoscopies is typically higher than in screening colonoscopies [4, 25]. Due to the high positive predictive value of the FOBT and FIT for CRC, the ASGE has raised the target ADR for this population over the screening population [26]. Also, we hypothesized that physician fatigue and procedure time would affect their ability to detect colorectal lesions. Generalized estimating equations (GEE) models to account for multiple polyps

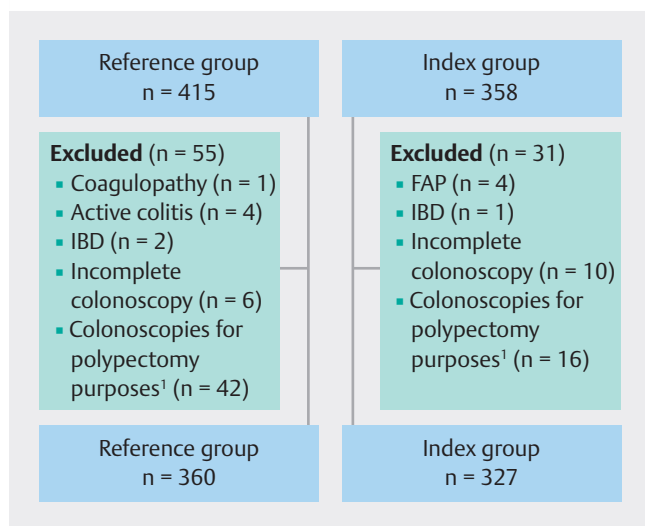


Fig. 1 Flowchart of the selection of patients. IBD, inflammatory bowel disease; FAP, familial adenomatous polyposis. ¹Colonoscopies for polypectomy purposes: endoscopic mucosal resection, endoscopic submucosal dissection, suspected positron emission tomography scan or virtual colonoscopy.

per patient, using a binomial distribution and a logarithm link function and generalized linear regressions were developed to compare the outcomes of interest, and the results were expressed as the relative risk (RR) and Beta (B) coefficients for linear regressions along with the 95% confidence intervals (Cis).

To estimate the random effect of endoscopists on the quality metrics of interest, the IPTW was estimated without endoscopists as a risk factor. The mixed models were created when endoscopists and IPTW were considered as random and fixed effects, respectively.

A sensitivity analysis compared the quality metrics between indexes and referents with positive and negative FIT results.

Sample size estimation can be found in the **Supplementary Material**. Significance was established at the 0.05 level (two-sided) in all comparisons. Analyses were performed using SPSS version 26.0 (IBM Corp., Armonk, New York, United States).

Results

Patient and procedure characteristics

A total of 687 patients were included (327 indexes and 360 referents) in the final analysis. **Fig. 1** shows the flowchart of patient selection. The majority of colonoscopies were performed for screening or adenoma surveillance. **Table 2** shows detailed patient and procedure characteristics.

Detection rates and mean number of detected polyps

In the reference and index cohorts, 355 and 453 polyps, respectively, were identified and underwent attempted resection (**Table 2**).

The PDR was higher in the index group (62.4%) than in the reference group (53.1%; $P=0.02$) (**Table 3**). The endoscopists

Table 2 Patient and procedure characteristics.

Variables	Reference group n = 360 (52.4%)	Index group n = 327 (47.6%)	P value
Age, median (IQR), years	64.0 (14.0)	65.0 (13.0)	0.25
Sex			0.09
▪ Male	170 (47.2)	176 (53.8)	
▪ Female	190 (52.8)	151 (46.2)	
First-degree family history of CRC, n (%) [†]			0.02
▪ No	239 (66.4)	235 (72.3)	
▪ Yes	77 (21.4)	70 (21.5)	
Unknown	44 (12.2)	20 (6.2)	
ASA classification, n (%) [†]			0.09
▪ I	110 (30.6)	105 (32.7)	
▪ II	229 (63.6)	208 (64.8)	
▪ III	21 (5.8)	8 (2.5)	
Colonoscopy indications, n (%)			0.03
▪ Screening	90 (25.0)	51 (15.6)	
▪ FIT [‡]	20 (5.6)	29 (8.9)	
▪ Adenoma surveillance	144 (40.0)	152 (46.5)	
▪ CRC surveillance	7 (1.9)	8 (2.4)	
▪ Anemia/bleeding	54 (15.0)	45 (13.8)	
▪ Diarrhea	15 (4.2)	9 (2.8)	
▪ Other	30 (8.3)	33 (10.1)	
Number of procedures performed by each endoscopist, n (%)			<0.001
▪ Endoscopist 1	129 (35.8)	170 (52.0)	
▪ Endoscopist 2	13 (3.6)	37 (11.3)	
▪ Endoscopist 3	140 (38.9)	70 (21.4)	
▪ Endoscopist 4	78 (21.7)	50 (15.3)	
Time of procedure, n (%)			0.54
▪ Morning	183 (50.8)	158 (48.3)	
▪ Afternoon	177 (49.2)	169 (51.7)	
Bowel preparation, n (%) [‡]			0.51
▪ Inadequate	36 (10.0)	27 (8.3)	
▪ Adequate	324 (90.0)	299 (91.7)	
Normal colonoscopies, n (%)	169 (46.9)	123 (37.6)	0.02
Colonoscopies with identified lesions, n (%)	191 (53.1)	204 (62.4)	
Lesions detected, n	355	453	

► **Table 2** (Continuation)

Variables	Reference group n = 360 (52.4%)	Index group n = 327 (47.6%)	P value
Anatomical segment, n (%) [§]			0.01
▪ Cecum	30 (8.5)	39 (8.6)	
▪ Ascending	49 (13.8)	96 (21.2)	
▪ Hepatic flexure	9 (2.5)	17 (3.8)	
▪ Splenic flexure	4 (1.1)	3 (0.7)	
▪ Transverse	66 (18.6)	91 (20.1)	
▪ Descending	49 (13.8)	82 (18.1)	
▪ Rectum	56 (15.8)	44 (9.7)	
▪ Sigmoid	89 (25.1)	80 (17.7)	
Paris classification, n (%) [¶]			0.18
▪ IS	120 (33.8)	356 (78.6)	
▪ IP	7 (2.0)	35 (7.7)	
▪ ISP	4 (1.1)	–	
▪ IIA	30 (8.5)	43 (9.5)	
▪ IIB	–	11 (2.4)	
Resection tool, n (%) ^{**}			<0.001
▪ Hot snare	26 (9.7)	32 (7.1)	
▪ Cold snare	208 (77.9)	312 (69.2)	
▪ Cold forceps	19 (7.1)	86 (19.1)	
▪ Other	12 (4.5)	2 (0.4)	
Resected, n (%)	347 (97.7)	432 (95.4)	0.09
Retrieved, n (%)	338 (95.2)	416 (91.8)	0.07
Pathology, n (%) ^{††}			<0.001
▪ Normal mucosa	1 (0.3)	36 (8.0)	
▪ Hyperplastic	98 (27.6)	98 (21.7)	
▪ Tubular	131 (36.9)	215 (47.6)	
▪ Tubulovillous	9 (2.5)	16 (3.5)	
▪ Villous	1 (0.3)	3 (0.7)	
▪ Traditional serrated	1 (0.3)	2 (0.4)	
▪ Sessile serrated lesions	19 (5.4)	28 (6.2)	
▪ High-grade dysplasia	11 (3.1)	3 (0.7)	
▪ Cancer	1 (0.3)	–	
▪ Other	59 (16.6)	15 (3.3)	
Polyp size, median (IQR), mm ^{‡‡}	4.0 (3.0)	3.0 (3.0)	<0.001

IQR, interquartile range; CRC, colorectal cancer; ASA, American Society of Anesthesiologists; FIT, fecal immunohistochemical test.

[§]Missing cases = 2 (0.3%).

[¶]Missing cases = 6 (0.9%).

[‡]Defined as total Boston Bowel Preparation Scale score <6; missing cases = 1 (0.3%).

[§]Missing controls = 3 (0.8%); missing cases = 1 (0.2%).

[¶]Missing controls = 194 (54.6%); missing cases = 8 (1.8%).

^{**}Missing cases = 19 (4.2%); missing controls = 2 (0.7%).

^{††}Missing cases = 36 (8.0%); missing controls = 24 (6.8%).

^{‡‡}Missing cases = 2 (0.4); missing controls = 17 (4.8%).

were about 50% (RR 1.47, 95% CI 1.08–1.99) more likely to detect at least one polyp during colonoscopies in the index group than in the reference group. When adjusted for confounders, the probability of polyp detection was 20% higher in the index group (RR 1.21, 95% CI 0.86–1.71).

The ADR was 11.1 percentage points higher in the index group than in the reference group (39.4% vs. 28.3%). The unadjusted and adjusted RRs were 1.65 (95% CI 1.20–2.27) and 1.51 (95% CI 1.05–2.17), respectively, indicating significant effect of a research assistant documenting the procedural characteristics on the ADR. Although the ADRs for the individual endoscopists increased in the index group, the difference reached the significance level only for one endoscopist (► **Table 3**).

Although the SLDR was higher in the index group compared to the reference group, the observed difference was not statistically significant (7.3% vs. 4.4%; $P=0.14$). After adjusting for confounders, the index group demonstrated a more than two-fold higher probability of detecting adenomatous serrated lesions compared to the reference group (RR 2.17, 95% CI 1.05–4.47) (► **Table 3**).

Moreover, the AADR was slightly but not statistically significantly higher in the index group, but this difference was reduced after adjusting for confounding factors.

The mean number of adenomatous polyps detected per patient (MAP) and (MASP) were significantly increased in the index group compared with the reference group ($P<0.001$). The regression analyses for both metrics demonstrated the direct and significant association between the Hawthorne effect and elevated MAP/MASP (► **Table 3**).

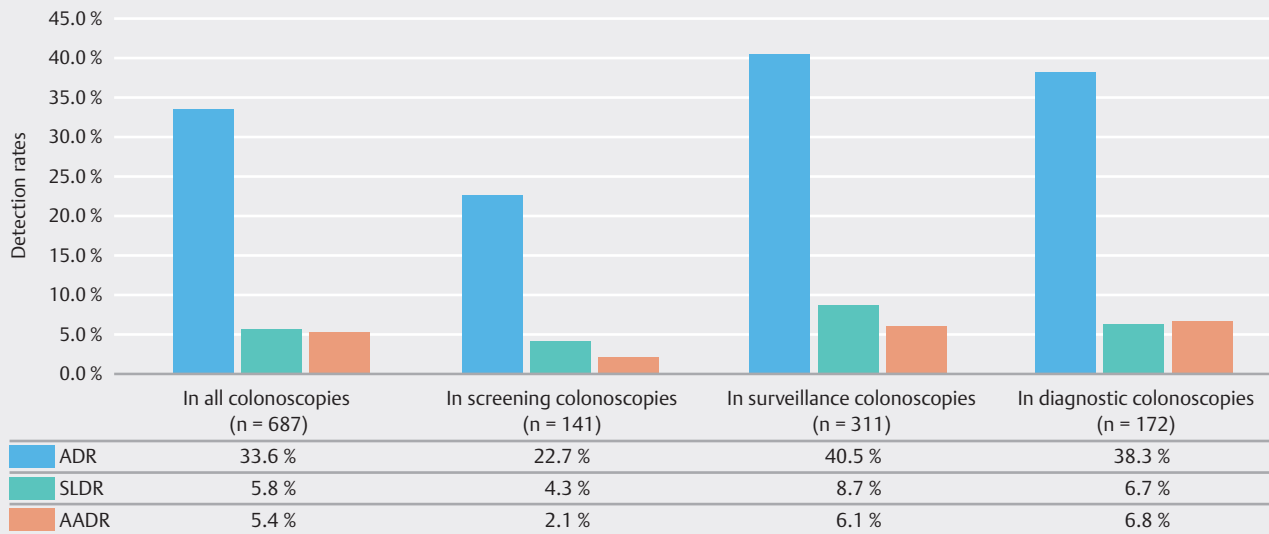
In contrast to detection rates, withdrawal time was shorter in the index group than in the reference group (6.9 vs. 8.0 minutes; $P=0.003$). Regression analyses showed an inverse relationship between the Hawthorne effect and withdrawal time; however, none of the point estimates reached statistical significance. Notably, withdrawal times in the index group were independently measured by the research assistant using a stopwatch shown on the recorded videos, while withdrawal times in the reference group were self-reported and obtained from the electronic health records, which makes them potentially inaccurate.

The mixed-effects models, with the endoscopist as a random effect and the group (index and reference groups) as fixed effects, yielded $P<0.001$ for all detection rates, MAP, and MASP. However, in the mixed model analysis, no significant variation in withdrawal time was observed among the endoscopists when comparing the reference group with the index group (P

► **Table 3** Comparison of colonoscopy quality metrics between indexes and referents.

Quality metric	Reference group n = 397 (54.8%)	Index group n = 327 (45.2%)	P value	Unadjusted RR (95%CI)	Adjusted RR (95%CI)
PDR, %	53.1	62.4	0.02	1.47 (1.08–1.99)	1.21 (0.86–1.71)
▪ Endoscopist 1	72.9	72.9	>0.99		
▪ Endoscopist 2	53.8	59.5	0.75		
▪ Endoscopist 3	40.7	52.9	0.11		
▪ Endoscopist 4	42.3	42.0	0.9		
ADR, %	28.3	39.4	0.002	1.65 (1.20–2.27)	1.51 (1.05–2.17)
▪ Endoscopist 1	24.8	41.2	0.003		
▪ Endoscopist 2	53.8	45.9	0.75		
▪ Endoscopist 3	28.6	37.1	0.21		
▪ Endoscopist 4	29.5	32.0	0.85		
SLDR, %	4.4	7.3	0.14	1.70 (0.89–3.27)	2.17 (1.05–4.47)
▪ Endoscopist 1	6.2	8.2	0.66		
▪ Endoscopist 2	7.7	0.0	0.26		
▪ Endoscopist 3	2.9	7.1	0.16		
▪ Endoscopist 4	3.8	10.0	0.26		
AADR, %	5.0	5.8	0.74	1.17 (0.60–2.27)	1.02 (0.47–2.22)
▪ Endoscopist 1	3.9	4.7	0.78		
▪ Endoscopist 2	23.1	10.8	0.36		
▪ Endoscopist 3	4.3	4.3	>0.99		
▪ Endoscopist 4	5.1	8.0	0.71		
MAP, mean (SD) ¹	0.5 (0.6)	0.8 (1.3)	<0.001	$\beta = 0.30 (0.15-0.46)$, $P < 0.0001$	$\beta = 0.21 (0.04-0.39)$, $P = 0.02$
▪ Endoscopist 1	0.4 (0.8)	0.7 (1.0)	0.003		
▪ Endoscopist 2	0.7 (0.9)	1.1 (1.8)	0.35		
▪ Endoscopist 3	0.4 (0.8)	0.7 (1.3)	0.07		
▪ Endoscopist 4	0.5 (0.9)	0.6 (1.1)	0.51		
MASP, mean (SD) ¹	0.5 (0.9)	0.8 (1.3)	<0.001	$\beta = 0.34 (0.17-0.50)$, $P < 0.0001$	$\beta = 0.26 (0.08-0.44)$, $P = 0.005$
▪ Endoscopist 1	0.5 (0.9)	0.8 (1.1)	0.005		
▪ Endoscopist 2	0.8 (0.9)	1.1 (1.8)	0.48		
▪ Endoscopist 3	0.5 (0.8)	0.8 (1.4)	0.03		
▪ Endoscopist 4	0.5 (0.9)	0.7 (1.2)	0.32		
Withdrawal time, median (IQR), minutes ^{1,2}	8.0 (2.0)	6.9 (4.7)	0.003	$\beta = -0.05 (-0.69-0.60)$, $P = 0.89$	$\beta = 0.08 (-0.70-0.85)$, $P = 0.85$
▪ Endoscopist 1	7.0 (6.0)	6.3 (4.3)	0.26		
▪ Endoscopist 2	12.0 (6.5)	6.9 (3.2)	<0.001		
▪ Endoscopist 3	8.0 (1.0)	7.7 (4.6)	0.001		
▪ Endoscopist 4	8.0 (2.0)	9.2 (8.1)	0.04		

RR, Relative risk; PDR, polyp detection rate; ADR, adenoma detection rate; SLDR, serrated lesion detection rate; AADR, advanced adenoma detection rate; MAP, mean number of adenomatous polyps detected per patient; MASP, mean number of adenomatous and sessile serrated polyps detected per patient; IQR, interquartile range; β , Beta coefficient. ¹Linear regression was used for the continuous variables. ²Missing cases = 8 (2.4%); missing controls = 21 (5.8%).



► **Fig. 2** Detection rates in all colonoscopies and in screening, diagnostic, and surveillance colonoscopies. ADR, adenoma detection rate; SLDR, serrated lesion detection rate; AADR, advanced adenoma detection rate.

= 0.26), as the residual variance for the individual endoscopists was negligible.

► **Table 3** shows details of the quality metrics in total and for individual endoscopists. ► **Fig. 2** displays the detection rates for all colonoscopies as well as for screening, surveillance, and diagnostic colonoscopies.

Sensitivity analysis

In the cohort of patients with a negative FIT result (n = 638), the PDR, ADR, and SLDR were higher in the index group. The association with the Hawthorne effect remained statistically significant in both unadjusted and adjusted regression analyses (► **Table 4**). The SLDR was increased by more than twofold in the index group compared with the reference group (adjusted RR 2.39, 95% CI 1.14–5.03). The AADR showed no statistically significant difference between the index and reference groups. However, the MAP and MASP showed a significant increase in the index group compared to the reference group. These associations remained significant after adjusting for confounders. The change in withdrawal time was similar to that for the whole cohort of patients. The variability of all detection rates, and the MAP/MASP rates among the endoscopists were significant in the mixed model analyses ($P < 0.001$), but not for the withdrawal time ($P = 0.29$) (► **Table 4**).

In contrast, in the cohort of patients with a positive FIT result (n = 49), no statistically significant differences were observed in detection rates, MAP, and MASP between the reference and index groups. Furthermore, regression analyses did not yield significant results after adjusting for confounders. Withdrawal time in the index group was approximately half that of the reference group ($P < 0.001$) and was inversely related to the Hawthorne effect. The random effects of endoscopists were significant for all metrics.

Discussion

In this study, we found that the presence of an observer significantly increased ADR, PDR, MAP and MASP during screening and surveillance colonoscopies. When comparing the index and reference groups, we observed an 11-percentage-point difference in the ADR. These increases in colonoscopy quality metrics remained significant after adjusting for potential confounders. Notably, the individual endoscopist also significantly influenced the outcomes. The validity of these findings was confirmed in the negative FIT cohort, with higher PDR and ADR observed in FIT-positive patients.

The ADR for all colonoscopies combined was 33.6% and did not vary among endoscopists (endoscopists 1 = 34.1%, 2 = 48.0%, 3 = 31.4%, 4 = 30.5% in the complete cohort). All endoscopists surpassed the recommended ADR threshold, with a significant increase in the index group [2], (► **Fig. 2**, ► **Fig. 3**). ADRs in screening colonoscopies (22.7%) were 10.9 percentage points lower than all colonoscopies for any indication, and lower than surveillance and diagnostic colonoscopies [25].

The existing evidence on the impact of the Hawthorne effect on ADR has been inconclusive and primarily focused on the presence of trainees, nurses, or fellows in the endoscopy room. A meta-analysis of five randomized controlled trials (RCTs) showed a statistically significant increase in ADR when two observers were present during endoscopies (33.9% vs. 29.5%; RR 1.24) [27]. In addition, the PDR was significantly higher for the observer group in four RCTs (43.3% vs. 40.0%; RR 1.31) [27]. In contrast, another meta-analysis of 14 studies found no difference between the ADR (RR 1.04, 95%CI 0.94–1.15) and PDR (RR 1.03, 95%CI 0.93–1.14) of colonoscopies with and without the attendance of fellows [28]. Neither study employed an independent observer who documented quality metrics and recorded full-length procedures without being di-

► **Table 4** Comparison of colonoscopy quality metrics between indexes and referents in patients with negative and positive fecal immunochemical test results.

Quality metric	Negative FIT				Positive FIT					
	Referents n = 340 (53.3%)	Indexes n = 298 (46.7%)	P value	Unadjusted RR (95%CI)	Adjusted RR (95%CI)	Referents n = 20 (40.8%)	Indexes n = 29 (59.2%)	P value	Unadjusted RR (95%CI)	Adjusted RR (95%CI)
PDR, %	51.5%	61.1%	0.02	1.48 (1.08–2.03)	1.26 (0.88–1.80)	80.0%	75.9%	>0.99	0.79 (0.20–3.15)	0.71 (0.18–2.88)
▪ Endoscopist 1	72.7	71.9	0.90			100	82.4	>0.99		
▪ Endoscopist 2	45.5	55.2	0.73			100	75.0	>0.99		
▪ Endoscopist 3	38.0	53.6	0.04			72.7	0.0	0.33		
▪ Endoscopist 4	38.9	40.4	>0.99			83.3	66.7	>0.99		
ADR, %	27.1	37.6	0.005	1.62 (1.16–2.27)	1.53 (1.05–2.24)	50.0	58.6	0.57	1.42 (0.45–4.46)	1.23 (0.39–3.99)
▪ Endoscopist 1	25.0	39.2	0.02			0.0	58.8	0.44		
▪ Endoscopist 2	45.5	41.4	>0.99			100	62.5	>0.99		
▪ Endoscopist 3	27.9	37.7	0.20			36.4	0.0	>0.99		
▪ Endoscopist 4	26.4	29.8	0.68			66.7	66.7	>0.99		
SLDR, %	4.1	8.1	0.04	2.04 (1.04–4.02)	2.39 (1.14–5.03)	10.0	0	0.16	-	-
▪ Endoscopist 1	6.3	9.2	0.50			-	-	-		
▪ Endoscopist 2	9.1	0.0	0.28			-	-	-		
▪ Endoscopist 3	2.3	7.2	0.13			9.1	0.0	>0.99		
▪ Endoscopist 4	2.8	10.6	0.11			16.7	0.0	>0.99		
AADR, %	4.4	4.7	>0.99	1.07 (0.51–2.25)	1.08 (0.46–2.54)	15.0	17.2	>0.99	1.18 (0.25–5.62)	0.99 (0.19–4.99)
▪ Endoscopist 1	3.9	3.9	>0.99			0.0	11.8	>0.99		
▪ Endoscopist 2	27.3	3.4	0.06			0.0	37.5	>0.99		
▪ Endoscopist 3	3.1	4.3	0.70			18.2	0.0	>0.99		
▪ Endoscopist 4	4.2	8.5	0.43			16.7	0.0	>0.99		
MAP, mean (SD) ¹	0.4 (0.8)	0.7 (1.1)	<0.001	$\beta = 0.27$, (-0.12–0.43), P<0.0001	$\beta = 0.21$ (-0.04–0.39), P=0.02	1.0 (1.2)	1.3 (2.0)	0.42	$\beta = 0.39$ (-0.47–1.25), P=0.37	$\beta = 0.09$ (-0.62–0.79), P=0.81
▪ Endoscopist 1	0.4 (0.8)	0.6 (1.0)	0.02			-	1.2 (1.4)	0.40		

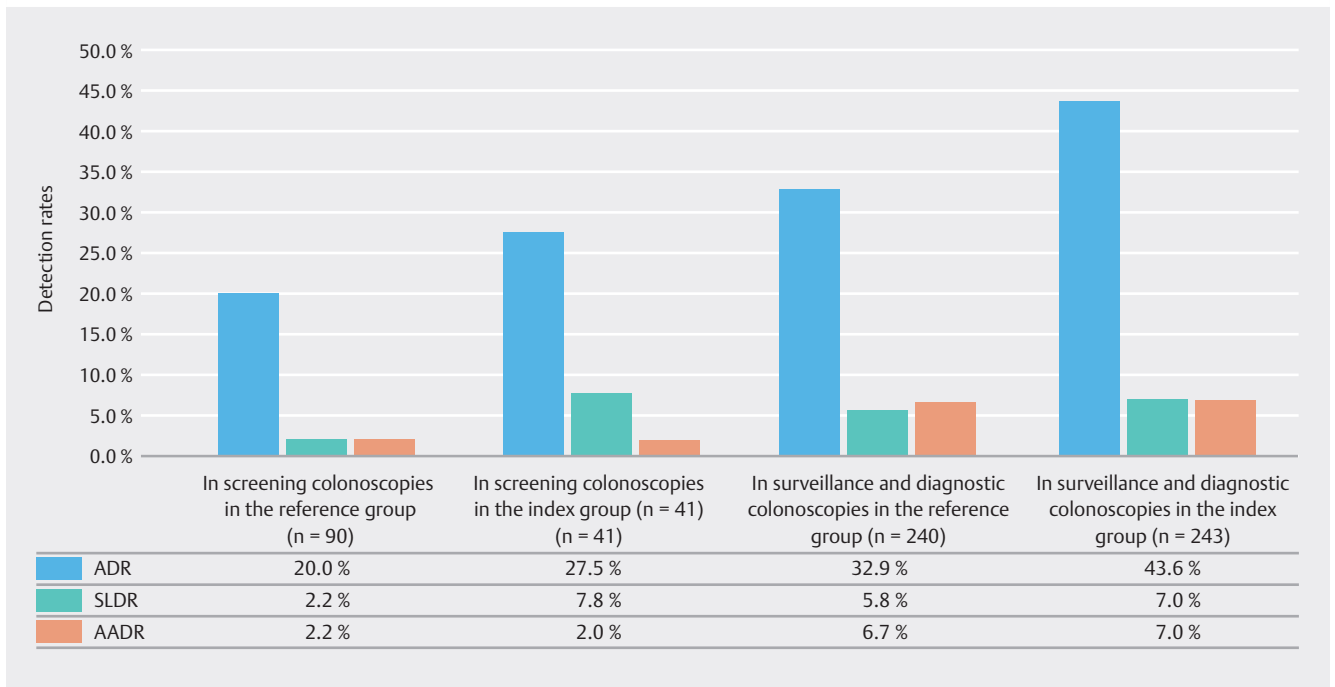
▶ **Table 4** (Continuation)

Quality metric	Negative FIT				Positive FIT					
	Referents n = 340 (53.3%)	Indexes n = 298 (46.7%)	P value	Unadjusted RR (95%CI)	Adjusted RR (95%CI)	Referents n = 20 (40.8%)	Indexes n = 29 (59.2%)	P value	Unadjusted RR (95%CI)	Adjusted RR (95%CI)
▪ Endoscopist 2	0.6 (0.9)	0.8 (1.3)	0.71			–	2.0 (3.1)	0.68		
▪ Endoscopist 3	0.4 (0.8)	0.8 (1.3)	0.02			0.7 (1.1)	–	0.54		
▪ Endoscopist 4	0.4 (0.7)	0.6 (1.1)	0.24			1.5 (1.5)	0.7 (0.6)	0.40		
MASP, mean (SD)*	0.5 (0.8)	0.8 (1.2)	<0.001	$\beta = 0.32$ (-0.16-0.48), P<0.0001	$\beta = 0.27$ (-0.08-0.45), P=0.006	1.1 (1.1)	1.3 (2.0)	0.56	$\beta = 0.30$ (-0.56-1.15), P=0.50	$\beta = -0.01$ (-0.70-0.68), P=0.97
▪ Endoscopist 1	0.5 (0.9)	0.7 (1.0)	0.02			–	1.2 (1.4)	0.40		
▪ Endoscopist 2	0.7 (1.0)	0.8 (1.3)	0.88			1.0 (0.0)	2.0 (3.1)	0.68		
▪ Endoscopist 3	0.4 (0.8)	0.9 (1.4)	0.006			0.8 (1.1)	–	0.48		
▪ Endoscopist 4	0.4 (0.7)	0.7 (1.2)	0.10			1.7 (1.4)	0.7 (0.6)	0.28		
Withdrawal time, median (IQR), minutes†	8.0 (1.0)	7.0 (4.8)	<0.001	$\beta = 0.30$ (-0.39-0.98), P = 0.40	$\beta = 0.30$ (-0.50-1.10), P = 0.47	10 (4.0)	5.6 (3.7)	<0.001	$\beta = -4.20$ (-5.80 to -2.60), P<0.001	$\beta = -4.17$ (-5.80 to -2.53), P<0.001
▪ Endoscopist 1	7.0 (2.0)	6.4 (4.1)	0.13			–	4.4 (2.9)	>0.99		
▪ Endoscopist 2	11.0 (7.0)	6.8 (3.8)	0.005			12.0 (0)	7.3 (1.9)	0.44		
▪ Endoscopist 3	8.0 (1.0)	7.7 (4.6)	0.001			8.0 (2.0)	–	>0.99		
▪ Endoscopist 4	8.0 (1.8)	9.2 (8.2)	0.01			10.0 (6.0)	8.4 (–)	0.5		

FIT, fecal immunologic test; RR, relative risk; PDR, polyp detection rate; ADR, adenoma detection rate; SADR, serrated lesion detection rate; AADR, advanced adenoma detection rate; MAP, mean number of adenomatous polyps detected per patient; MASP, mean number of adenomatous and sessile serrated polyps detected per patient; IQR, interquartile range.

*Linear regression was used for the continuous variables.

†Missing cases = 8 (2.7%); missing controls = 20 (5.9)



► **Fig. 3** Detection rates in screening colonoscopies and in surveillance and diagnostic colonoscopies in the index and reference groups. ADR, adenoma detection rate; SLDR, serrated lesion detection rate; AADR, advanced adenoma detection rate.

rectly involved in the medical procedure. We believe that our study design accurately reflects an independent observation effect on increasing endoscopist vigilance without involvement in the clinical process, providing a better representation of the Hawthorne effect. Noteworthy, the utilization of video recording in conjunction with the presence of a research assistant may have had a synergistic effect on the observed Hawthorne effect. However, due to the retrospective design of this study, it was not possible to quantify the extent of this effect. To address this issue, additional prospective or randomized studies are needed to provide more definitive insights.

Previous studies did not report or adjust ADR for colonoscopy indications or individual endoscopist performance. We attempted to control for confounding patient- and procedure-related factors. Interestingly, our findings highlight the significant impact of individual endoscopists on the association between colonoscopy quality metrics and the Hawthorne effect after accounting for this influence in the adjusted analysis. Unlike the study mentioned above, we did not observe an increase in withdrawal time due to the Hawthorne effect. This difference may be due to the self-reporting of withdrawal time in the reference group, while the index group had their withdrawal times measured by a research assistant using a stopwatch. This may lead to more precise documentation in the index group, as endoscopists in the reference group might have rounded up the time to the nearest whole minute. One endoscopist had a longer withdrawal time, but in the mixed model analysis, the variability between endoscopists did not reach significance, suggesting minimal impact of endoscopists on this quality metric.

Recent publications have primarily focused on integrating advanced devices or imaging technology for improving colo-

noscopy ADR [5]. Our study reveals that independent observer documentation improves ADR, highlighting quality monitoring impact on colonoscopy. Many centers now monitor and provide feedback on ADR to endoscopists. Moreover, our study highlights the need to consider the potential impact of the Hawthorne effect when interpreting ADR reports in studies evaluating AI-assisted systems, where the presence of a research assistant during the procedure was neglected. It seems that apart from AI's computational power in detecting colorectal lesions, the vigilance of endoscopists using such systems may also contribute to AI's superior outcome. Therefore, our study shows that an independent observer might be a simple solution for increasing ADR or a significant contributor to increased detection when implementing AI-based colonoscopy solutions [2]. In the future, AI systems are expected to act as a second observer in colonoscopy procedures, alerting endoscopists to the presence of polyps and automating the reporting of important components like withdrawal time and detection of landmarks [8].

The association between ADR and SLDR has been well-described in previous research [29, 30, 31]. The prevalence of serrated lesions in our study (6.2% in the whole cohort) was slightly lower than the prevalence reported in other studies [29, 30]. Interestingly, significant correlations were observed between PDR and SLDR ($P < 0.001$), but no significant correlation was found between ADR and SLDR in the entire cohort and the cohort of patients with FIT-negative and -positive results ($P=0.85$, 0.50, and 0.11, respectively). Our findings differ from other studies that have concluded that FIT-based screening programs may not effectively detect advanced serrated lesions, as these lesions are less likely to produce a positive FIT result. Given that the minimal ADR threshold for FIT-positive population is

higher than the acceptable threshold for the screening population, we found that the PDR, ADR, AADR, MAP, and MASP of patients with a positive FIT result were significantly higher than in patients with a negative FIT result [26, 32]. However, documentation of the procedure by an independent observer did not improve these metrics in FIT-positive patients. Due to the small sample size, we could not assess the potential influence of the Hawthorne effect on the SLDR. Notably, we stratified the patients based on FIT results and included all colonoscopies performed for any indication in the comparison groups. All colonoscopies were also performed by the same endoscopists, eliminating the random effect of endoscopists on the polyp detection.

Some limitations must be acknowledged for this study. There is a possibility of selection bias due to the retrospective design of the reference group and the different processes involved in the recruitment of index and referent patients. We attempted to address this bias by applying the same exclusion criteria to the reference group as used in the index group, and by adjusting for the other confounding factors. We focused solely on the performance of expert gastroenterologists in our study and did not assess the influence of factors such as training year, expertise level, or training program (gastroenterology or surgery) on detection rates and other quality metrics by including fellows. Therefore, generalization of the results to other endoscopists with different backgrounds or training standards must be done with caution. The likelihood of the increase in the PDR, AADR, and withdrawal time due to the Hawthorne effect did not reach statistical significance after adjusting for potential confounding factors; however, this may have been due to insufficient study power.

Conclusions

In conclusion, our study highlights the significant impact of the Hawthorne effect on colonoscopy quality metrics. These findings support the consideration of the Hawthorne effect, presence of an assistant during colonoscopy procedures, and recording of the procedures, as a potential contributor to improving quality in clinical practice and when interpreting research results evaluating ADR. Although this adjustment may involve additional costs, it is a feasible solution that can lead to better detection rates, improved procedure quality, and favorable patient outcomes.

Conflict of Interest

Daniel vonRenteln is supported by a "Fonds de Recherche du Québec Santé" career development award. He has also received research funding from ERBE, Ventage, Pendopharm, and Pentax, and is a consultant for Boston Scientific and Pendopharm. The remaining authors declare that they have no conflict of interest.

References

- [1] 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: 687–696 doi:10.1056/NEJMoa1100370
- [2] Rex DK, Schoenfeld PS, Cohen J et al. Quality indicators for colonoscopy. *Gastrointest Endosc* 2015; 81: 31–53 doi:10.1016/j.gie.2014.07.058
- [3] Cross AJ, Robbins EC, Saunders BP et al. Higher adenoma detection rates at screening associated with lower long-term colorectal cancer incidence and mortality. *Clinical Gastroenterology and Hepatology* 2022; 20: e148–e167
- [4] Kahi CJ, Vemulapalli KC, Johnson CS et al. Improving measurement of the adenoma detection rate and adenoma per colonoscopy quality metric. *The Indiana University experience* 2014; 79: 448–454
- [5] Taghiakbari M, Mori Y, von Renteln D. Artificial intelligence-assisted colonoscopy: A review of current state of practice and research. *World J Gastroenterol* 2021; 27: 8103–8122 doi:10.3748/wjg.v27.i47.8103
- [6] Aniwan S, Orkoonsawat P, Viriyautsahakul V et al. The secondary quality indicator to improve prediction of adenoma miss rate apart from adenoma detection rate. *Am J Gastroenterol* 2016; 111: 723–729
- [7] Srigley JA, Furness CD, Baker GR et al. Quantification of the Hawthorne effect in hand hygiene compliance monitoring using an electronic monitoring system: A retrospective cohort study. *BMJ Qual Saf* 2014; 23: 974–980 doi:10.1136/bmjqs-2014-003080
- [8] Taghiakbari M, Hamidi Ghalehjegh S, Jehanno E et al. Automated detection of anatomical landmarks during colonoscopy using a deep learning model. *J Canadian Assoc Gastroenterol* 2023; doi:10.1093/jcag/gwad017
- [9] Kaltenbach T, Anderson JC, Burke CA et al. Endoscopic removal of colorectal lesions: Recommendations by the US multi-society task force on colorectal cancer. *Am J Gastroenterol* 2020; 115: 435–464 doi:10.14309/ajg.0000000000000555
- [10] Jass JR, Sobin LH, Morson BC. *Histological typing of intestinal tumours*. 2nd ed. Berlin, New York: Springer-Verlag; 1989
- [11] 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 doi:10.14309/ajg.0000000000000544
- [12] The Paris endoscopic classification of superficial neoplastic lesions. Esophagus, stomach, and colon: November 30 to December 1, 2002. *Gastrointest Endosc* 2003; 58: S3–S43 doi:10.1016/s0016-5107(03)02159-x
- [13] Wang H, Wang P, Liu X et al. Factors predicting the colorectal adenoma detection rate in colonoscopic screening of a Chinese population: A prospective study. *Medicine (Baltimore)* 2019; 98: e15103 doi:10.1097/MD.00000000000015103
- [14] Bailie L, Loughrey MB, Coleman HGJG. Lifestyle risk factors for serrated colorectal polyps: A systematic review and meta-analysis. *Gastroenterology* 2017; 152: 92–104 doi:10.1053/j.gastro.2016.09.003
- [15] Kolligs FT, Crispin A, Munte A et al. Risk of advanced colorectal neoplasia according to age and gender. *PLoS One* 2011; 6: e20076 doi:10.1371/journal.pone.0020076
- [16] Lin J, O'Connor E, Evans C et al. Screening for Colorectal Cancer: An Evidence Update for the U.S. Preventive Services Task Force [Internet]. 2021: <https://www.ncbi.nlm.nih.gov/books/NBK570913/>
- [17] Rex DK, Johnson DA, Anderson JC et al. American College of Gastroenterology guidelines for colorectal cancer screening 2008. *Am J Gastroenterol* 2009; 104: 739–750

- [18] Kaminski MF, Regula J, Kraszewska E et al. Quality indicators for colonoscopy and the risk of interval cancer. *N Engl J Med* 2010; 362: 1795–1803 doi:10.1056/NEJMoa0907667
- [19] Regula J, Rupinski M, Kraszewska E et al. Colonoscopy in colorectal-cancer screening for detection of advanced neoplasia. *N Engl J Med* 2006; 355: 1863–1872 doi:10.1056/NEJMoa054967
- [20] Brenner H, Altenhofen L, Hoffmeister MJA. Sex, age, and birth cohort effects in colorectal neoplasms: A cohort analysis. *Ann Intern Med* 2010; 152: 697–703 doi:10.7326/0003-4819-152-11-201006010-00002
- [21] Ferlitsch M, Reinhart K, Pramhas S et al. Sex-specific prevalence of adenomas, advanced adenomas, and colorectal cancer in individuals undergoing screening colonoscopy. *JAMA* 2011; 306: 1352–1358 doi:10.1001/jama.2011.1362
- [22] Ricci E, Hassan C, Petruzzello L et al. Inter-centre variability of the adenoma detection rate: A prospective, multicentre study. *Dig Liver Dis* 2013; 45: 1022–1027 doi:10.1016/j.dld.2013.05.009
- [23] Anderson JC, Butterly LF, Goodrich M et al. Differences in detection rates of adenomas and serrated polyps in screening versus surveillance colonoscopies, based on the new hampshire colonoscopy registry. *Clin Gastroenterol Hepatol* 2013; 11: 1308–1312 doi:10.1016/j.cgh.2013.04.042
- [24] Millan MS, Gross P, Manilich E et al. Adenoma detection rate: The real indicator of quality in colonoscopy. *Dis Colon Rectum* 2008; 51: 1217–1220 doi:10.1007/s10350-008-9315-3
- [25] Hassan C, Repici A, Rex DK. Fitting adr to colonoscopy indication. *United European Gastroenterol J* 2017; 5: 149–152 doi:10.1177/2050640616667171
- [26] Rembacken B, Hassan C, Riemann JF et al. Quality in screening colonoscopy: Position statement of the European Society of Gastrointestinal Endoscopy (ESGE). *Endoscopy* 2012; 44: 957–968 doi:10.1055/s-0032-1325686
- [27] Aziz M, Weissman S, Khan Z et al. Use of 2 observers increases adenoma detection rate during colonoscopy: Systematic review and meta-analysis. *Clin Gastroenterol Hepatol* 2020; 18: 1240–1242
- [28] Oh YS, Collins CL, Virani S et al. Lack of impact on polyp detection by fellow involvement during colonoscopy: A meta-analysis. *Digest Dis Sci* 2013; 58: 3413–3421
- [29] Anderson JC, Butterly LF, Weiss JE et al. Providing data for serrated polyp detection rate benchmarks: An analysis of the new hampshire colonoscopy registry. *Gastrointest Endosc* 2017; 85: 1188–1194 doi:10.1016/j.gie.2017.01.020
- [30] Hazewinkel Y, de Wijkerslooth TR, Stoop EM et al. Prevalence of serrated polyps and association with synchronous advanced neoplasia in screening colonoscopy. *Endoscopy* 2014; 46: 219–224
- [31] Schramm C, Janhsen K, Hofer J-H et al. Detection of clinically relevant serrated polyps during screening colonoscopy: Results from seven cooperating centers within the german colorectal screening program. *Endoscopy* 2018; 50: 993–1000
- [32] Cubiella J, Castells A, Andreu M et al. Correlation between adenoma detection rate in colonoscopy- and fecal immunochemical testing-based colorectal cancer screening programs. *United European Gastroenterol J* 2017; 5: 255–260