Journal of Pathology
J Pathol 2024
Published online 29 March 2024 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/path.6276

ORIGINAL ARTICLE

Polyketide synthase positive Escherichia coli one-time measurement in stool is not informative of colorectal cancer risk in a screening setting

Willemijn de Klaver[1,2], Meike de Wit[2], Anne Bolijn[2], Marianne Tijssen[2], Pien Delis-van Diemen[2], Margriet Lemmens[2], Manon CW Spaander[3], Evelien Dekker[1], Monique E van Leerdam[4,5], Veerle MH Coupé[6], Ruben van Boxtel[7,8], Hans Clevers[7,8,9,10,11], Beatriz Carvalho[2] and Gerrit A Meijer[2*]

[1] *Department of Gastroenterology and Hepatology, Amsterdam University Medical Centers, Location University of Amsterdam, Amsterdam, The Netherlands*
[2] *Department of Pathology, Netherlands Cancer Institute, Amsterdam, The Netherlands*
[3] *Department of Gastroenterology and Hepatology, Erasmus University Medical Center, Rotterdam, The Netherlands*
[4] *Department of Gastrointestinal Oncology, Netherlands Cancer Institute, Amsterdam, The Netherlands*
[5] *Department of Gastroenterology and Hepatology, Leiden University Medical Center, Leiden, The Netherlands*
[6] *Department of Epidemiology and Data Science, Amsterdam University Medical Centers, Location VU Medical Center, Amsterdam, The Netherlands*
[7] *Princess Máxima Center for pediatric oncology, Utrecht, The Netherlands*
[8] *Oncode Institute, Utrecht, The Netherlands*
[9] *University Medical Center Utrecht, Utrecht, The Netherlands*
[10] *Hubrecht Institute, Utrecht, the Netherlands*
[11] *Pharma, Research and Early Development (pRED) of F. Hoffmann-La Roche Ltd, Basel, Switzerland*

**Correspondence to: GA Meijer, Department of Pathology, Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX, Amsterdam, The Netherlands. E-mail: g.meijer@nki.nl*

Abstract

Environmental factors like the pathogenicity island *polyketide synthase positive (pks+) Escherichia coli* (*E. coli*) could have potential for risk stratification in colorectal cancer (CRC) screening. The association between *pks+ E. coli* measured in fecal immunochemical test (FIT) samples and the detection of advanced neoplasia (AN) at colonoscopy was investigated. Biobanked FIT samples were analyzed for both presence of *E. coli* and *pks+ E. coli* and correlated with colonoscopy findings; 5020 CRC screening participants were included. Controls were participants in which no relevant lesion was detected because of FIT-negative results (cut-off ≥15 μg Hb/g feces), a negative colonoscopy, or a colonoscopy during which only a nonadvanced polyp was detected. Cases were participants with AN [CRC, advanced adenoma (AA), or advanced serrated polyp (ASP)]. Existing DNA isolation and quantitative polymerase chain reaction (qPCR) procedures were used for the detection of *E. coli* and *pks+ E. coli* in stool. A total of 4542 (90.2%) individuals were *E. coli* positive, and 1322 (26.2%) were *pks+ E. coli* positive. The prevalence of *E. coli* in FIT samples from individuals with AN was 92.9% compared to 89.7% in FIT samples of controls ($p = 0.010$). The prevalence of *pks+ E. coli* in FIT samples from individuals with AN (28.6%) and controls (25.9%) was not significantly different ($p = 0.13$). The prevalences of *pks+ E. coli* in FIT samples from individuals with CRC, AA, or ASP were 29.6%, 28.3%, and 32.1%, respectively. In conclusion, the prevalence of *pks+ E. coli* in a screening population was 26.2% and did not differ significantly between individuals with AN and controls. These findings disqualify the straightforward option of using a snapshot measurement of *pks+ E. coli* in FIT samples as a stratification biomarker for CRC risk.

Keywords: colorectal cancer; advanced neoplasia; screening; *pks+ E. coli*; risk stratification; biomarker

Received 21 August 2023; Revised 21 December 2023; Accepted 22 February 2024

Conflict of interest statement: MdW has several patents pending and/or issued. MdW is also cofounder, stockholder and board member (COO) of CRCbioscreen BV. MCWS has received research support from Medtronic, Boston Scientific, Sentinel, Sysmex and Norgine. ED has endoscopic equipment on loan from FujiFilm and has received a research grant from FujiFilm. ED has received honoraria for consultancy from Fujifilm, Olympus, InterVenn, Ambu and Exact Sciences and speakers' fees from Olympus, GI Supply, Norgine, IPSEN, PAION and FujiFilm. ED is a supervisory board member of the eNose Company. ED is also Global Chair of the CRC Screening Committee of the World Endoscopy Organisation, Chair of the Dutch Post-polypectomy surveillance Guideline and member of the Post-polypectomy surveillance guideline of the European Society of Gastrointestinal Endoscopy. VC and BC have several patents pending and/or issued. HC's full disclosure is given at https://www.uu.nl/staff/JCClevers/. HC is inventor on several patents related to organoid technology and currently an employee of Roche, Basel. GAM is cofounder, stockholder and board member (CSO) of CRCbioscreen BV, has a research collaboration with CZ Health Insurances (cash matching to ZonMW grant) and research collaborations with Exact Sciences, Sysmex, Sentinel Ch. SpA, Personal Genome Diagnostics (PGDX), DELFi and Hartwig Medical Foundation; these companies provide

materials, equipment and/or sample/genomic analyses. GAM is also CSO of Health-RI (Dutch National Health Data infrastructure for research and innovation) and member of the supervisory board of IKNL (Netherlands Comprehensive Cancer Organisation). GAM also has several patents pending and/or issued. WdK received consulting fees as member of the Dutch Post-polypectomy surveillance guideline committee. AB, MT, PDD, ML, MEvL and RvB declared no competing interests.

Introduction

The risk of developing colorectal cancer (CRC) is determined by genetic and environmental factors, the latter including human gut microbiota and their metabolites [1–5]. Multiple bacterial species have been associated with CRC, but demonstrating causal relationships has remained challenging [6]. A particular strain of the common gut bacterium *Escherichia coli* (*E. coli*) can be regarded as an exception, however [7]. The genome of this strain contains the pathogenicity island polyketide synthase (*pks*), a gene cluster that encodes for colibactin biosynthetic enzymes [7,8]. Colibactin can induce interstrand DNA cross-linking, which can lead to double-strand DNA breaks, chromosome aberrations, and cell-cycle arrest [8–17]. Repeated exposure of normal human colonic epithelial organoids to *pks*+ *E. coli* caused a specific *pks*+ mutational signature, as measured by whole genome sequencing (WGS) [7]. This signature is present in genes commonly mutated in CRC, like the *adenomatous polyposis coli* (*APC*) gene [6,7,18]. In addition, the signature has been found in ∼8% of 496 CRCs analyzed by WGS that are present in the Hartwig Medical Foundation database, which stores genetic and clinical data of metastatic cancers diagnosed in the Netherlands [7]. These findings provided important arguments for considering *pks*+ *E. coli* as an environmental risk factor for CRC.

Early detection of CRC is the most effective approach to reduce CRC-related morbidity and mortality, and CRC screening in average-risk populations has been demonstrated to be successful in many countries [19,20]. Yet, there is a persisting need to optimize the benefit-to-harm ratio of CRC screening, in particular to prevent overdiagnosis [21,22].

Many ongoing CRC screening programs use the fecal immunochemical test (FIT), which measures human hemoglobin (Hb) in stool, to identify individuals at risk of CRC and refer them for colonoscopy. The predictive value of FIT-based CRC screening depends on the test features sensitivity and specificity, but also on the pretest likelihood of presence of the disease, that is, the prevalence of advanced neoplasia [advanced neoplasia (AN), CRC, and/or advanced precursor lesions] in the population screened. To increase the positive predictive value of FIT, most CRC screening programs only use age as CRC risk factor by screening individuals in specific age categories [23]. However, better identification of individuals at risk of CRC could be useful to improve the benefit-to-harm ratio of CRC screening. Therefore, there is an evident need for objective and quantifiable risk factors to personalize CRC early detection. This study therefore aimed to explore the potential clinical utility of *pks*+ *E. coli* status as an environmental risk factor for CRC to optimize the predictive value of FIT-based CRC screening. To this end, the association between *pks*+ *E. coli* measured in FIT samples and the detection of AN at colonoscopy was evaluated.

Materials and methods

Study population

Biobanked FIT samples from individuals participating in one of two previously performed CRC screening trials, i.e. the COlonoscopy or COlonography for Screening (COCOS) trial and the Fecal Immunochemical Test comparison (FIT comparison) trial, were used [24,25]. The COCOS trial was executed before the start of the Dutch national CRC screening program, between 2009 and 2010. In this trial, colonoscopy and noncathartic computed tomography (CT) colonography were evaluated as primary screening tests, while FIT samples (OC sensor, Eiken Chemical Co., Tokyo, Japan) were collected. FIT samples were included only from participants that underwent colonoscopy. Individuals with AN were included as cases and individuals without AN/any neoplasia at colonoscopy as controls. The FIT comparison trial, which was conducted in the context of the Dutch national CRC screening program between 2016 and 2017, compared the diagnostic performance of two different FIT brands (OC Sensor, Eiken Chemical Co., Tokyo, Japan, and FOB Gold, Sentinel, Milan, Italy). Both FIT samples were collected from the same bowel movement, and if one or both FITs were positive (cut-off ≥15 μg/g feces), individuals were referred for colonoscopy. Individuals with AN were included as cases and individuals without AN/any neoplasia at colonoscopy or who did not undergo a colonoscopy due to two negative FITs were included as controls. Written informed consent was obtained from all participants in both trials. Ethical approval, including biobanking for further research, for the COCOS trial was obtained from the Dutch Minister of Health, Welfare and Sport (2009/03WBO, The Hague, The Netherlands), and the study was registered in the Dutch Trial Registry: NTR1829. Ethical approval for the FIT comparison trial, including biobanking for further research, was also obtained from the Dutch Minister of Health, Welfare and Sport (Population Screening Act; no. 769500-1357 16-PG), and the trial was also registered in the Dutch Trial Registry: NTR5874. Ethical approval for reuse of the fecal samples for this study was obtained from

the Institutional Review Board of the Netherlands Cancer Institute (IRBdm20-096) in 2020.

Public and patient representatives were involved through the Dutch Digestive Foundation and the Dutch Cancer Society.

Stool samples consisted of leftover material in FIT collection devices. All colonoscopies were performed by gastroenterologists with a lifetime experience of 500 or more colonoscopies. All lesions detected and resected were evaluated by experienced pathologists. Individuals were classified based on their most advanced lesion. An advanced adenoma (AA) was defined as an adenoma ≥10 mm, and/or with a ≥25% villous component, and/or with high-grade dysplasia. An advanced serrated polyp (ASP) was defined as a serrated polyp ≥10 mm and/or with any grade of dysplasia. The category AN included CRC, AA, and ASP. Controls were defined as individuals in which no relevant lesion was detected, i.e. individuals without AN at colonoscopy (COCOS and FIT comparison trials) and individuals that did not undergo colonoscopy because of two negative FITs (cut-off ≥15 μg Hb/g feces) (FIT comparison trial).

To determine the prevalence of *pks+ E. coli* in the average-risk population a series of samples of adequate size were randomly selected, with both FIT positive as well as negative test results. In cross-sectional studies of patient populations, it was found that prevalence of *pks+ E. coli* bacterium can reach up to 20%. Therefore, we performed a sample size calculation under the assumption of a prevalence of 20% in the population and the requirement to reach an accuracy of this prevalence estimate of 2.25% for the width of the 95% confidence interval (CI). Using the normal approximation for CI determination for proportions, this resulted in the requirement to include 5000 individuals in this study.

OC-Sensor FIT samples from 1040 participants of the COCOS trial were used. All participants were screening naïve and between 50 and 75 years of age. The most advanced lesions in this population were CRC ($n = 8$), AA ($n = 93$), ASP ($n = 28$), nonadvanced adenoma ($n = 202$), and nonadvanced serrated polyp ($n = 129$). In total, 580 participants had a negative colonoscopy (no colorectal neoplasia was detected).

FOB-Gold FIT samples from 3980 participants of the FIT comparison trial were used, unless these were not available, in which case OC Sensor FIT samples were used. All participants were screening naïve and between 59 and 75 years of age. The most advanced lesions in this population were CRC ($n = 73$), AA ($n = 486$), nonadvanced adenoma ($n = 394$), and serrated polyp unspecified ($n = 74$). In total 412 participants had a negative colonoscopy (no colorectal neoplasia was detected), and 2541 participants did not undergo colonoscopy because they were FIT-negative.

FIT analysis for CRC screening and storage

At the time of the COCOS trial, FIT samples were first collected and stored at −80 °C before they were analyzed in batches, within a number of weeks, to obtain a quantitative FIT result. At the time of the FIT comparison trial, FIT samples were analyzed upon arrival in the laboratory to obtain a quantitative FIT result. The OC Sensor samples were analyzed using the OC Sensor DIANA automated analyzer (Eiken Chemical), and the FOB-Gold samples were analyzed using the BioMajesty JCA-BM6010/C analyzer (Jeol Diagnostic Systems, Beijing, PR China). After analysis, FIT collection tubes with leftover material were stored at −80 °C in a well-annotated biobank.

Characteristics of study individuals

In total, 5020 biobanked FIT samples were used for this study. Of those, 1056 (21.0%) were OC Sensor and 3964 (79.0%) were FOB Gold. The 5020 individuals had the following characteristics: 2702 (53.8%) were male, median age was 60 years (IQR 58–63), and 1150 individuals (22.9%) were FIT-positive (cut-off ≥15 μg Hb/g feces) based on the FIT sample used for the *E. coli* and *pks+ E. coli* analysis. Of all individuals, 688 (13.7%) had AN, and in 4258 (84.8%) individuals no relevant lesion was detected, of whom 992 (19.8%) had a negative colonoscopy during which no colorectal neoplasia was detected, and 2541 (50.6%) had two negative FIT results (cut-off ≥15 μg Hb/g feces) and did not undergo colonoscopy (Table 1).

DNA extraction from FIT samples

For this study, DNA was extracted from a 200-μl aliquot of leftover FIT samples using the ZymoBIOMICS 96 DNA kit using Lysis Tubes (Zymo Research, Irvine, CA, USA) following the manufacturer's instructions. This kit allows for high-throughput DNA isolation using 96-well plates. The isolated DNA was immediately used for quantitative polymerase chain reaction (qPCR) analysis.

Detection of (*pks+*) *E. coli* in FIT samples by qPCR

First, an evaluation was conducted to determine whether *pks+ E. coli* could be reliably measured in FIT samples. A qPCR targeting *clbB* was performed to detect *pks+* bacteria [26–28]. *E. coli* was used to demonstrate the presence of *E. coli* bacteria (Table 2). Existing qPCR protocols for *pks+ E. coli* detection in (whole) stool samples were used, but to cope with the lesser input material in FIT samples, the input DNA was maximized (5 μl undiluted FIT DNA) [27,28]. qPCR reactions were run in duplicate by performing a multiplex PCR on 5 μl isolated DNA using the Taqman Gene Expression Mastermix (Thermo Fisher Scientific, Waltham, MA, USA) supplemented with BSA (0.1 μg/μl) on a Quantstudio 6 Flex System (Thermo Fisher Scientific). The cycling conditions for *E. coli* and *clbB* were as follows: 2 min at 50 °C for optimal removal of carryover contaminants, an initial denaturation step of 10 min at 95 °C, followed by 40 cycles of 15 s at 95 °C and 1 min at 60 °C. A *pks+* bacterial DNA sample (CCR20, kindly provided by the Hubrecht Institute, Utrecht, the Netherlands) was taken along as positive control. In every qPCR plate, a standard

Table 1. Baseline characteristics of the study population.

	Total population (n = 5020, 100.0%)	COCOS trial population (n = 1040, 20.7%)	FIT comparison trial population (n = 3980, 79.3%)
Demographics			
Male, *n* (%)	2702 (53.8%)	532 (51.2%)	2170 (54.5%)
Age, years (IQR)	60 (58–63)	60 (55–65)	60 (59–63)
Fecal immunochemical test screening			
OC Sensor	1056 (21.0%)	1040 (100.0%)	16 (0.4%)
FOB Gold	3964 (79.0%)	-	3964 (99.6%)
FIT-positive result (cut-off ≥15 µg Hb/g feces)	1150 (22.9%)	74 (7.1%)	1076 (27.0%)
Findings at screening colonoscopy			
Advanced neoplasia, *n* (%)	688 (13.7%)	129 (12.4%)	559 (14.0%)
Colorectal cancer	81 (1.6%)	8 (0.8%)	73 (1.8%)
Advanced adenoma	579 (11.5%)	93 (8.9%)	486 (12.2%)
Advanced serrated polyp	28 (0.6%)	28 (2.7%)	-
Nonadvanced adenoma, *n* (%)	596 (11.9%)	202 (19.4%)	394 (9.9%)
Serrated polyp, *n* (%)	203 (4.0%)	129 (12.4%)	74 (1.9%)
Nonadvanced serrated polyp	129 (2.6%)	129 (12.4%)	-
Serrated polyp unspecified*	74 (1.5%)	-	74 (1.9%)
Control groups, *n* (%)			
No relevant lesion detected†	4258 (84.8%)	911 (87.6%)	3347 (84.1%)
Negative colonoscopy‡	992 (19.8%)	580 (55.8%)	412 (10.4%)
FIT-negative result (cut-off ≥ 15 µg Hb/g feces)§	2541 (50.6%)	-	2541 (63.8%)

COCOS, COlonoscopy versus COlonography Screening; FIT, fecal immunochemical test; Hb, hemoglobin; IQR, interquartile range; -, not applicable.
*Serrated polyp unspecified = a serrated polyp whose size and in which the presence of dysplasia are unknown.
†No relevant lesion detected = a negative FIT result, a negative colonoscopy, or a colonoscopy during which only nonadvanced adenoma or nonadvanced serrated polyp were detected.
‡Negative colonoscopy = a colonoscopy during which no colorectal neoplasia was detected.
§FIT-negative result = no colonoscopy, but two fecal immunochemical tests (OC Sensor and FOB Gold) performed in the same bowel movement were both negative (cut-off ≥ 15 µg Hb/g feces).

Table 2. qPCR primers and probes.

Target		Sequence (5'–3')	Product size	Reference
Pks (clbB)	Forward	GCGCATCCTCAAGAGTAAATA	280 bp	[27]
	Reverse	GCGCTCTATGCTCATCAACC		
	Probe	5'FAM-TATTCGACACAGAACAACGCCGGT-3'BHQ1		[28]
E. coli	Forward	CATGCCGCGTGTATGAAGAA	96 bp	[27]
	Reverse	CGGGTAACGTCAATGAGCAAA		
	Probe	5'HEX-TCGGGTTGTAAAGTACTTTCAGCGGG-3'BHQ1		[28]

Bp, base pairs; *E. coli*, *Escherichia coli*; *pks*, polyketide synthase; qPCR, quantitative polymerase chain reaction.

curve, using this *pks*+ bacterial DNA, was incorporated in order to quantify the number of *pks*+ bacteria detected [29]. The mean of the duplicates was calculated and used as quantitative *E. coli* and *pks*+ *E. coli* results. The threshold used for detection was 10 copies/5 µl of DNA template. A *pks*− *E. coli* bacterial DNA sample (CFF22-1D5, also kindly provided by the Hubrecht Institute, Utrecht, the Netherlands) was used as a negative control [13]. The operators in the laboratory who performed the DNA isolations and qPCRs to obtain quantitative *E. coli* and *pks*+ *E. coli* results were blinded for the colonoscopy findings.

Statistical analysis

Qualitative data were reported as percentages, and nonnormally distributed data were reported as medians with interquartile ranges (IQRs). A Mann–Whitney *U*-test was used to compare continuous variables between groups for nonnormally distributed data. The χ^2 test was used for comparison of dichotomous variables between groups. Odds ratios (ORs) with a 95% CI were calculated for the associations between *E. coli* and *pks*+ *E. coli*, respectively, with the presence of AN. *P* values of <0.05 were considered statistically significant. All statistical analyses were performed in IBM SPSS Statistics (version 27; https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-27). GraphPad Prism (version 9; Graphpad Inc, San Diego, CA, USA) was used for all graphs.

Results

Detection of (*pks*+) *E. coli* in FIT samples by qPCR

Detection of *E. coli* and *pks*+ *E. coli* using qPCR was feasible even in the minute amounts of stool present in FIT samples after performing the original FIT analysis. Of the 5020 individuals tested, 4526 (90.2%, 95% CI: 89.3–91.0) tested positive for *E. coli* (at a cut-off of ≥10 copies/5 µl of DNA template). Of all individuals positive for *E. coli*, 1313 (29.0%) were also positive for *pks* (cut-off of ≥10 copies/5 µl of DNA template). Thus, in

Table 3. Prevalence of *E. coli* and *pks+ E. coli* in a screening population.

Most advanced lesion	Total population ($n = 5020$)
E. coli status*	
Positive, *n* (%)	4526 (90.2%)
Negative, *n* (%)	494 (9.8%)
Pks E. coli status*	
Positive, *n* (%)	1313 (26.2%)
Negative, *n* (%)	3707 (73.8%)

E. coli, Escherichia coli; *pks*, polyketide synthase.
*Positivity cut-off was ≥10 copies/5 µl of DNA template.

26.2% (95% CI: 24.9–27.4) of all average-risk screening participants ($n = 5020$), *pks+ E. coli* was detected in their FIT sample (Table 3). Of the 494 participants negative for *E. coli*, only six (1.2%) were positive for *pks*. The prevalence of *E. coli* and *pks+ E. coli* was not significantly different in the COCOS trial population compared to the FIT comparison trial population (supplementary material, Table S1). The prevalence of *E. coli* and *pks+ E. coli* presented by sex and age at FIT can be found in supplementary material, Table S2, where for sex, no statistically significant differences were found.

The prevalence of *E. coli* in individuals with AN and controls was 92.9% (95% CI: 90.7–94.7) and 89.7% (95% CI: 88.8–90.6) ($p = 0.010$), respectively. The prevalence of *E. coli* in individuals with CRC, AA, or ASP was 96.3% (95% CI: 89.6–99.2), 92.4% (95% CI: 89.9–94.4), and 92.9% (95% CI: 76.5–99.1), respectively. However, of all individuals with AN, only individuals with AA had a significantly different prevalence of *E. coli* compared to controls ($p = 0.043$). The prevalence of *pks+ E. coli* in individuals with AN, CRC, AA, or ASP or controls was 28.6% (95% CI: 25.3–32.2), 29.6% (95% CI: 20.0–40.8), 28.3% (95% CI: 24.7–32.2), 32.1% (95% CI: 15.9–52.4%), and 25.9% (95% CI: 24.6–27.3), respectively. There was no significant difference between the prevalence of *pks+ E. coli* in individuals with AN and controls ($p = 0.131$). Moreover, none of the differences observed in *pks+ E. coli* prevalence between individuals with CRC, AA, or ASP and controls were statistically significant (Table 4).

For individuals with nonadvanced adenomas, nonadvanced serrated polyps, and serrated polyps unspecified, the prevalence of *E. coli* was 89.6% (95% CI: 86.9–91.9), 88.4% (95% CI: 81.6–93.3), and 90.5% (95% CI: 81.5–96.1), respectively. The prevalence of *pks+ E. coli* for the same lesions was 25.8% (95% CI: 22.4–29.6), 27.1% (95% CI: 19.7–35.7), and 17.6% (95% CI: 9.79–28.2), respectively. Comparison of the prevalence of *E. coli* and *pks+ E. coli* for all lesion types to different control groups and age categories did not result in substantially different observations (supplementary material, Tables S3 and S4), nor was there a statistically significant difference in *E. coli* and *pks+ E. coli* prevalence for distally or proximally located AN (supplementary material, Table S5) [18].

In addition to identifying samples as positive or negative, using a fixed threshold (10 copies/5 µl of DNA template), we also quantified the *E. coli* and *pks*

Table 4. Prevalence of *E. coli* and *pks+ E. coli* in screened individuals with advanced neoplasia during colonoscopy compared with controls without relevant lesion detected.

Most advanced lesion	Total population ($n = 5020$)	Advanced neoplasia ($n = 688$)	Colorectal cancer ($n = 81$)	Advanced adenoma ($n = 579$)	Advanced serrated polyp ($n = 28$)	No relevant lesion detected ($n = 4258$)
E. coli status*						
Positive, *n* (%, 95 CI)	4526 (90.2, 89.3–91.0)	639 (92.9, 90.7–94.7)	78 (96.3, 89.6–99.2)	535 (92.4, 89.9–94.4)	26 (92.9, 76.5–99.1)	3820 (89.7, 88.8–90.6)
Negative, *n* (%, 95 CI)	494 (9.8, 9.0–10.7)	49 (7.1, 5.3–9.3)	3 (3.7, 0.8–10.4)	44 (7.6, 5.6–10.1)	2 (7.1, 0.9–23.5)	438 (10.3, 9.4–11.2)
E. coli status compared to control group in last column (*p* value)†						
No relevant lesion detected	-	0.010	0.052	0.043	0.585	-
Pks E. coli status*						
Positive, *n* (%, 95 CI)	1313 (26.2, 24.9–27.4)	197 (28.6, 25.3–32.2)	24 (29.6, 20.0–40.8)	164 (28.3, 24.7–32.2)	9 (32.1, 15.9–52.4)	1103 (25.9, 24.6–27.3)
Negative, *n* (%, 95 CI)	3707 (73.8, 72.6–75.1)	491 (71.4, 67.8–74.7)	57 (70.4, 59.2–80.0)	415 (71.7, 67.8–75.3)	19 (67.9, 47.7–84.1)	3155 (74.1, 72.8–75.4)
Pks E. coli status compared to control group in last column (*p* value)‡						
No relevant lesion detected	-	0.131	0.449	0.214	0.453	-

E. coli, Escherichia coli; *pks*, polyketide synthase. -, not applicable.
*Positivity cut-off is ≥10 copies/5 µl of DNA template.
†,‡χ^2 test.

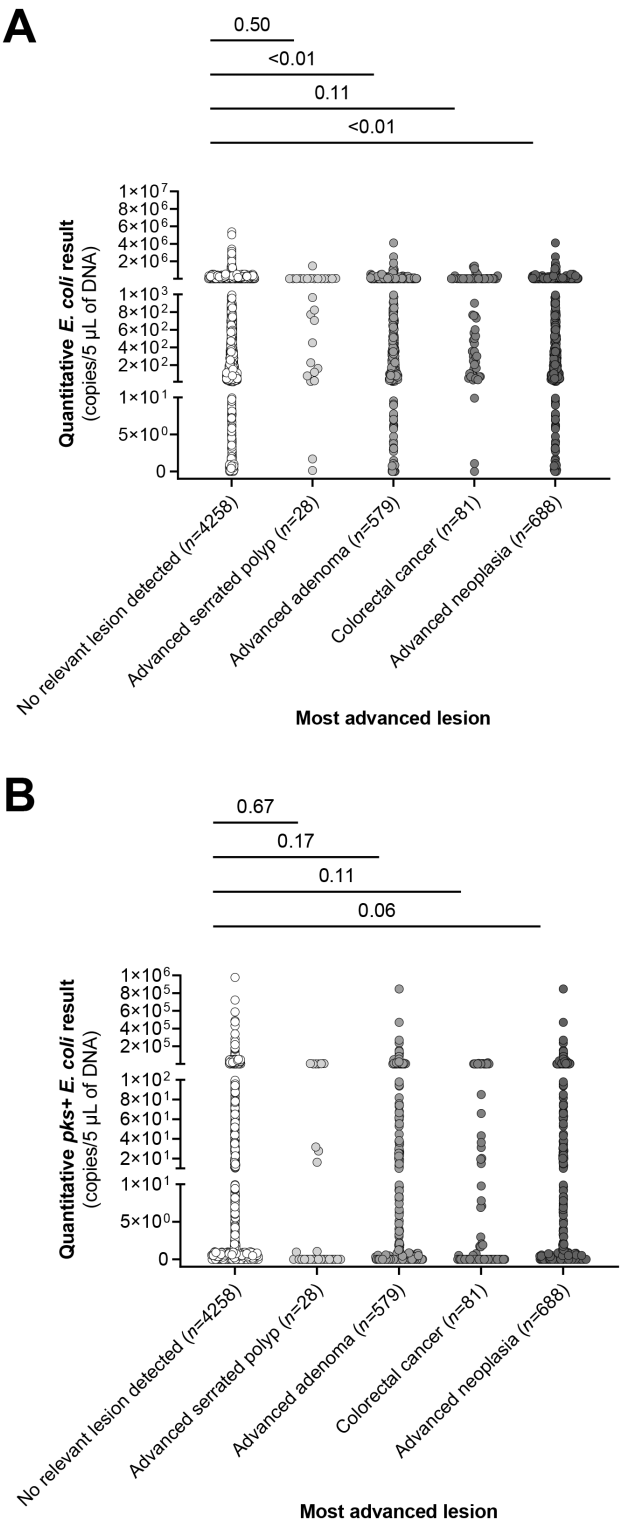

Figure 1. Quantitative *E. coli* and *pks+ E. coli* result per lesion type. (A) Quantification of *E. coli* in individuals with AN (CRC, AA, or ASP) versus individuals with no relevant lesion detected. (B) Quantification of *pks+ E. coli* in individuals with AN (CRC, AA, or ASP) versus individuals with no relevant lesion detected. A Mann–Whitney *U*-test was used to compare the continuous variables between groups. Figure created with GraphPad Prism.

+ *E. coli* results, i.e. the number of copies/5 µl of DNA template and evaluated this for individuals with AN, CRC, AA, and ASP and for controls. Again, a difference of the number of *E. coli* copies (mean rank) was observed for individuals with AN ($p < 0.01$), compared to controls. The number of *pks+ E. coli* copies (mean rank) did not differ significantly between individuals with AN and controls (Figure 1). Comparing the number of *E. coli* and *pks+ E. coli* copies in individuals with CRC, AA, and ASP to different control groups did not result in substantially different observations (supplementary material, Figure S1).

Discussion

Demonstrating causal relationships between bacterial species and CRC is challenging, especially due to the multifactorial nature of cancer development (1–6). *pks+ E. coli* is an exception in this respect as it was shown to induce specific DNA mutations that may lead to CRC. This study aimed to evaluate the most straightforward approach to using *pks+ E. coli* as an environmental risk factor in the context of FIT-based CRC screening, i.e. a snapshot-in-time approach. If the presence of *pks+ E. coli* were measured directly in FIT samples, the approach would be scalable and implementable in population-based CRC screening programs.

Indeed, *E. coli* and *pks+ E. coli* could be detected reliably in FIT samples using qPCR. Moreover, in the large average-risk CRC screening population of the more than 5000 individuals studied, 90.2% of individuals were *E. coli* positive and 26.2% were *pks+ E. coli* positive. These data provide the most precise estimate of *pks+ E. coli* prevalence in a CRC screening population available today. Earlier reports on the prevalence of *pks+ E. coli* are discordant and frequently based on small sample sizes [13,27,30–35]. In addition, most studies reporting on the prevalence of *pks+ E. coli* used mucosa-associated samples, e.g. fresh frozen tissue or formalin-fixed, paraffin-embedded samples, which may yield different results compared to stool samples, and this sampling approach is not scalable for routine use in CRC screening programs [13,27,34,35]. Importantly, in this study, the risk of finding AN during colonoscopy in a single screening round was not associated with the presence of *pks+ E. coli* in FIT samples. Consequently, these findings convincingly disqualify the straightforward option of using a snapshot measurement of *pks+ E. coli* in FIT samples as a stratification biomarker for CRC risk in CRC early detection.

Our findings are in line with those of two previously published Japanese studies that looked at the prevalence of *pks+ E. coli* in stool ($n = 968$) and colonic lavage ($n = 98$) samples [30,31]. However, a Swedish study ($n = 240$) and a Chinese study ($n = 139$) that also evaluated the prevalence of *pks+ E. coli* in stool both found a significant difference in *pks+ E. coli* prevalence in individuals with CRC compared to controls [32,33]. In addition, a recent meta-analysis on the association of *pks+ E. coli* and the development of CRC included 12 articles of which only two focused on *pks+ E. coli* detection in stool. The meta-analysis showed that individuals with *pks+ E. coli* seemed to have an increased

risk of developing CRC; however, these results were based mainly on studies that evaluated tissue samples, and at least four of the included articles used non-Western study populations [36]. As the studies published to date were small, used selected series, and showed inconsistent results, a large study within the intended-use population, like the present one, is essential for generating conclusive data for *pks+ E. coli* prevalence in the average-risk CRC screening population and for identifying an association of presence of *pks+ E. coli* with AN.

Interestingly, the prevalence of *E. coli* in individuals with AN (92.9%) was slightly, but significantly, higher ($p = 0.010$) than that in individuals in whom no relevant lesions were detected (89.7%). While the absence of an association between *pks+ E. coli* in stool and AN in the colon is somewhat disappointing from a clinical biomarker perspective, it does provide additional insights into the apparently complex interaction between *pks+ E. coli* and CRC. While the observations in a well-controlled organoid setting are evident, their translation to the clinical setting appears less straightforward. For *pks+ E. coli* to impose CRC risk, it likely must exert its effects on the colorectal epithelium over a long period of time, as do most environmental risk factors. Much remains unknown in terms of longitudinal infestation or interaction with other risk factors with this microorganism. One option is that an extrinsic and regionally acting mutagenic agent would colonize crypts in healthy individuals, causing mutations that might, later in life, result in disease [7,37–40]. In line with this, a prevalence of *pks+ E. coli* of 15–18% in newborns at 3 days of life has been observed, whereas the present study shows a prevalence of 26.2% in the CRC screening age range [39,40]. Yet little is known about the pattern of infestation in between these age ranges or about any variation in susceptibility to acquiring *pks+ E. coli*-induced mutations in the colon. Data from longitudinal birth cohorts may provide answers on this issue. In addition, *pks+ E. coli* has been shown to induce mutations in specific driver genes (e.g. *APC*) early in carcinogenesis [7]. Based on these findings, we hypothesized that cells carrying these mutations were primed for malignant transformation later in life, thereby requiring fewer additional (age-related) mutations to develop into CRC.

A strength of this large-scale study is that all FIT samples were retrieved from an average-risk screening population. In addition, stool samples were collected at the moment of intended use, prior to bowel preparation and colonoscopy, using a routine FIT collection device. Taken together, this study has high external validity. Nevertheless, some limitations should be addressed.

While this study had a large sample size of an average-risk screening-age population, the findings still represent a single geographic population. Whether the prevalence of *pks+ E. coli* and the lack of association with cancer risk are uniform across countries when analyzed at this scale remains to be determined. Furthermore, *pks* has also been shown to be present in other gut bacteria like *Klebsiella pneumoniae*, *Klebsiella aerogenes*, and *Citrobacter koseri* [16,41]. Indeed, 0.5% of *pks*-positive cases were *E. coli*-negative, while the approach used did not allow us to further determine which species would host *pks* in these cases. Assuming that *pks* in other bacteria species would also produce colibactin and yield the minute number of such cases observed, this is unlikely to have a significant impact on the findings of this study. In addition, ideally the load and duration of the *pks+ E. coli* exposure should be known to estimate the risk imposed. However, as individuals are unaware of their *pks+ E. coli* status over time and longitudinal measurements of this variable are not readily available, such data are not available. Alternatively, the impact of *pks+ E. coli* infestation could be deduced from the accumulation of *pks*-specific mutations, i.e. the *pks+* mutational signature, in colonic normal or adenoma epithelium. Yet, so far, this signature can only reliably be determined by WGS of DNA from fresh frozen tissue samples, while from the adenomas in this study only formalin-fixed, paraffin-embedded tissue samples are available. Moreover, such a WGS-based approach is unlikely to be cost effective, a crucial condition for adoption in population-based screening programs.

This study did not set out to determine the technically best approach for estimating the risk of CRC attributable to exposure to *pks+ E. coli* but rather aimed to investigate the pragmatic approach—whether measuring *pks+ E. coli* in routinely collected FIT samples could be informative of such a risk, which clearly was not the case. This does not preclude the possibility that gut microbiota, either fecal or mucosa-associated, may carry such a risk. However, sampling, for example, mucosa-associated bacteria from the large intestine is more complex than taking a FIT sample and likely less feasible from a population-based cancer screening perspective.

In addition to *pks+ E. coli*, several other gram-negative bacterial toxins (e.g. cytolethal distending toxin, cytotoxic necrotizing factor, cycle inhibiting factor, and bacteroides fragilis toxin) have been associated with CRC. Nevertheless, larger studies based on samples from CRC screening populations should be performed to evaluate their potential for risk stratification in CRC screening. Moreover, the challenges associated with using bacterial status in stool at a single point in time as a stratification biomarker for CRC risk may well apply to bacterial toxins in general [42,43].

In conclusion, *pks+ E. coli* is common in adults at screening age. No association was found between *pks+ E. coli* status in stool at a single point in time and findings at colonoscopy, rendering this approach infeasible for the risk stratification of participants in FIT-based CRC screening programs. Longitudinal studies will be required to further elucidate the impact of timing, frequency, and duration of *pks+ E. coli* exposure on future CRC-risk.

Disclaimer

The funding sources had no role in the design of this study or during its execution, analyses, data interpretation, or decision to submit results.

Acknowledgements

We would like to thank all individuals who participated in the COCOS and FIT trials and all researchers that made these trials possible. Support for this study was provided both by the Dutch Digestive Foundation (WOO 20-02 and project OPSLAG) and by a Stand Up to Cancer/Dutch Cancer Society International Translational Cancer Research Dream Team Grant (SU2C–AACR–DT1415, MEDOCC). Stand Up to Cancer is a program of the Entertainment Industry Foundation administered by the American Association for Cancer Research.

Author contribution statement

GAM, ED, VMHC, BC, MdW and WdK conceived and designed the study. MT and AB performed all laboratory analysis. WdK and MdW performed analyses. GAM, BC, MdW and WdK drafted the manuscript. All authors (WdK, MdW, AB, MT, PDD, ML, MCWS, ED, MEvL, VMHC, RvB, HC, BC and GAM) have critically evaluated the content of the manuscript and agreed with its submission.

Data availability statement

Data are available upon reasonable request.

References

1. Tjalsma H, Boleij A, Marchesi JR, *et al*. A bacterial driver-passenger model for colorectal cancer: beyond the usual suspects. *Nat Rev Microbiol* 2012; **10:** 575–582.
2. O'Keefe SJ. Diet, microorganisms and their metabolites, and colon cancer. *Nat Rev Gastroenterol Hepatol* 2016; **13:** 691–706.
3. Wan M-L, Wang Y, Zeng Z, *et al*. Colorectal cancer (CRC) as a multifactorial disease and its causal correlations with multiple signaling pathways. *Biosci Rep* 2020; **40:** BSR20200265.
4. Hanus M, Parada-Venegas D, Landskron G, *et al*. Immune system, microbiota, and microbial metabolites: the unresolved triad in colorectal cancer microenvironment. *Front Immunol* 2021; **12:** 612826.
5. Jain T, Sharma P, Are AC, *et al*. New insights into the cancer-microbiome-immune axis: decrypting a decade of discoveries. *Front Immunol* 2021; **12:** 622064.
6. Rosendahl Huber A, Pleguezuelos-Manzano C, Puschhof J. A bacterial mutational footprint in colorectal cancer genomes. *Br J Cancer* 2021; **124:** 1751–1753.
7. Pleguezuelos-Manzano C, Puschhof J, Rosendahl Huber A, *et al*. Mutational signature in colorectal cancer caused by genotoxic pks$^+$ *E. coli*. *Nature* 2020; **580:** 269–273.
8. Nougayrède JP, Homburg S, Taieb F, *et al*. *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. *Science* 2006; **313:** 848–851.
9. Bossuet-Greif N, Vignard J, Taieb F, *et al*. The colibactin genotoxin generates DNA Interstrand cross-links in infected cells. *mBio* 2018; **9:** e02393-17.
10. Xue M, Kim CS, Healy AR, *et al*. Structure elucidation of colibactin and its DNA cross-links. *Science* 2019; **365:** eaax2685.
11. Wilson MR, Jiang Y, Villalta PW, *et al*. The human gut bacterial genotoxin colibactin alkylates DNA. *Science* 2019; **363:** eaar7785.
12. Taieb F, Nougayrède JP, Watrin C, *et al*. *Escherichia coli* cyclomodulin Cif induces G2 arrest of the host cell cycle without activation of the DNA-damage checkpoint-signalling pathway. *Cell Microbiol* 2006; **8:** 1910–1921.
13. Buc E, Dubois D, Sauvanet P, *et al*. High prevalence of mucosa-associated *E. coli* producing cyclomodulin and genotoxin in colon cancer. *PLoS One* 2013; **8:** e56964.
14. Nougayrède JP, Taieb F, De Rycke J, *et al*. Cyclomoduflins: bacterial effectors that modulate the eukaryotic cell cycle. *Trends Microbiol* 2005; **13:** 103–110.
15. Cuevas-Ramos G, Petit CR, Marcq I, *et al*. *Escherichia coli* induces DNA damage in vivo and triggers genomic instability in mammalian cells. *Proc Natl Acad Sci U S A* 2010; **107:** 11537–11542.
16. Faïs T, Delmas J, Barnich N, *et al*. Colibactin: more than a new bacterial toxin. *Toxins (Basel)* 2018; **10:** 151.
17. Dziubańska-Kusibab PJ, Berger H, Battistini F, *et al*. Colibactin DNA-damage signature indicates mutational impact in colorectal cancer. *Nat Med* 2020; **26:** 1063–1069.
18. Terlouw D, Suerink M, Boot A, *et al*. Recurrent APC splice variant c.835-8A>G in patients with unexplained colorectal polyposis fulfilling the colibactin mutational signature. *Gastroenterology* 2020; **159:** 1612–1614.e5.
19. Sung H, Ferlay J, Siegel RL, *et al*. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021; **71:** 209–249.
20. Ladabaum U, Dominitz JA, Kahi C, *et al*. Strategies for colorectal cancer screening. *Gastroenterology* 2020; **158:** 418–432.
21. Knudsen AB, Zauber AG, Rutter CM, *et al*. Estimation of benefits, burden, and harms of colorectal cancer screening strategies: modeling study for the US preventive services task force. *JAMA* 2016; **315:** 2595–2609.
22. Gray JA, Patnick J, Blanks RG. Maximising benefit and minimising harm of screening. *BMJ* 2008; **336:** 480–483.
23. Navarro M, Nicolas A, Ferrandez A, *et al*. Colorectal cancer population screening programs worldwide in 2016: an update. *World J Gastroenterol* 2017; **23:** 3632–3642.
24. Wieten E, de Klerk CM, van der Steen A, *et al*. Equivalent accuracy of 2 quantitative fecal immunochemical tests in detecting advanced neoplasia in an organized colorectal cancer screening program. *Gastroenterology* 2018; **155:** 1392–1399.e5.
25. Stoop EM, de Haan MC, de Wijkerslooth TR, *et al*. Participation and yield of colonoscopy versus non-cathartic CT colonography in population-based screening for colorectal cancer: a randomised controlled trial. *Lancet Oncol* 2012; **13:** 55–64.
26. Johnson JR, Johnston B, Kuskowski MA, *et al*. Molecular epidemiology and phylogenetic distribution of the *Escherichia coli* pks genomic Island. *J Clin Microbiol* 2008; **46:** 3906–3911.
27. Arthur JC, Perez-Chanona E, Mühlbauer M, *et al*. Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science* 2012; **338:** 120–123.
28. Drewes JL, Corona A, Sanchez U, *et al*. Transmission and clearance of potential procarcinogenic bacteria during fecal microbiota transplantation for recurrent *Clostridioides difficile*. *JCI Insight* 2019; **4:** e130848.
29. Cougnoux A, Dalmasso G, Martinez R, *et al*. Bacterial genotoxin colibactin promotes colon tumour growth by inducing a senescence-associated secretory phenotype. *Gut* 2014; **63:** 1932–1942.
30. Shimpoh T, Hirata Y, Ihara S, *et al*. Prevalence of pks-positive *Escherichia coli* in Japanese patients with or without colorectal cancer. *Gut Pathog* 2017; **9:** 35.
31. Iwasaki M, Kanehara R, Yamaji T, *et al*. Association of *Escherichia coli* containing polyketide synthase in the gut microbiota with colorectal neoplasia in Japan. *Cancer Sci* 2022; **113:** 277–286.

32. Liu K, Yang X, Zeng M, *et al*. The role of fecal *Fusobacterium nucleatum* and pks$^+$ *Escherichia coli* as early diagnostic markers of colorectal cancer. *Dis Markers* 2021; **2021:** 1171239.
33. Eklöf V, Löfgren-Burström A, Zingmark C, *et al*. Cancer-associated fecal microbial markers in colorectal cancer detection. *Int J Cancer* 2017; **141:** 2528–2536.
34. Iyadorai T, Mariappan V, Vellasamy KM, *et al*. Prevalence and association of pks+ *Escherichia coli* with colorectal cancer in patients at the university Malaya medical Centre, Malaysia. *PLoS One* 2020; **15:** e0228217.
35. Villariba-Tolentino C, Cariño AM, Notarte KI, *et al*. pks$^+$ *Escherichia coli* more prevalent in benign than malignant colorectal tumors. *Mol Biol Rep* 2021; **48:** 5451–5458.
36. Gaab ME, Lozano PO, Ibañez D, *et al*. A meta-analysis on the association of colibactin-producing pks+ *Escherichia coli* with the development of colorectal cancer. *Lab Med* 2023; **54:** 75–82.
37. Lee-Six H, Olafsson S, Ellis P, *et al*. The landscape of somatic mutation in normal colorectal epithelial cells. *Nature* 2019; **574:** 532–537.
38. Berger H, Meyer TF. Mechanistic dissection unmasks colibactin as a prevalent mutagenic driver of cancer. *Cancer Cell* 2021; **39:** 1439–1441.
39. Nowrouzian FL, Oswald E. *Escherichia coli* strains with the capacity for long-term persistence in the bowel microbiota carry the potentially genotoxic pks Island. *Microb Pathog* 2012; **53:** 180–182.
40. Payros D, Secher T, Boury M, *et al*. Maternally acquired genotoxic *Escherichia coli* alters offspring's intestinal homeostasis. *Gut Microbes* 2014; **5:** 313–325.
41. Morgan RN, Saleh SE, Farrag HA, *et al*. Prevalence and pathologic effects of colibactin and cytotoxic necrotizing factor-1 (Cnf 1) in *Escherichia coli*: experimental and bioinformatics analyses. *Gut Pathog* 2019; **11:** 22.
42. Piciocchi A, Germinario EAP, Garcia Etxebarria K, *et al*. Association of polygenic risk score and bacterial toxins at screening colonoscopy with colorectal cancer progression: a multicenter case-control study. *Toxins (Basel)* 2021; **13:** 569.
43. Mezerova K, Raclavsky V, Stary L. Which bacterial toxins are worthy of validation as markers in colorectal cancer screening? A critical review. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub* 2022; **166:** 1–11.

SUPPLEMENTARY MATERIAL ONLINE

Figure S1. The number of *E. coli* and *pks+ E. coli* copies in individuals with CRC, AA, and ASP compared to different control groups

Table S1. Prevalence of *E. coli* and *pks+ E. coli* in a screening population presented separately for each screening trial

Table S2. Prevalence of *E. coli* and *pks+ E. coli* in a screening population presented by sex and age

Table S3. Prevalence of *E. coli* and *pks+ E. coli* in a screening population per lesion type and compared to different control groups

Table S4. Prevalence of *E. coli* and *pks+ E. coli* in a screening population per lesion type and different control groups presented by age category

Table S5. Prevalence of *E. coli* and *pks+ E. coli* in those screened from the COCOS trial with advanced neoplasia during colonoscopy presented by lesion location (proximal versus distal)