From the DEPARTMENT OF CLINICAL SCIENCE AND
EDUCATION, SÖDERSJUKHUSET
Karolinska Institutet, Stockholm, Sweden

FECAL IMMUNOCHEMICAL TEST IN COLORECTAL CANCER SCREENING
-impact of screening strategy and gender on colonoscopy findings, missed lesions and costs

Hanna Ribbing Wilén

Karolinska Institutet

Stockholm 2022

Fecal Immunochemical Test in colorectal cancer screening
-impact of screening strategy and gender on colonoscopy findings, missed lesions and costs

THESIS FOR DOCTORAL DEGREE (Ph.D)

By

Hanna Ribbing Wilén

The thesis will be defended in public at Södersjukhuset, sal Ihre, Friday 8th of April, 2022, at 9:00 am

Principal Supervisor:
Ass. Professor Johannes Blom
Karolinska Institutet
Department of Clinical Science and Education, Södersjukhuset
Division of Surgery

Co-supervisor(s):
M.D., Ph.D. Deborah Saraste
Karolinska Institutet
Department of Clinical Science and Education, Södersjukhuset
Division of Surgery

Ass. Professor Folke Hammarqvist
Karolinska Institutet
Department of Clinical Science, Intervention and Technology
Division of Surgery

Opponent:
Professor Michal Kaminski
Maria-Sklodowska-Curie Memorial Cancer Center and Institute of Oncology
Department of Cancer Prevention and Department of Gastroenterological Oncology
Warsaw, Poland

Examination Board:
Ass. Professor Birger Pålsson
Lund University, Faculty of Medicine
Department of Clinical Sciences, Lund
Division of Surgery

Ass. Professor Annika Sjövall
Karolinska Institutet
Department of Molecular Medicine and Surgery
Division of Colorectal Surgery

Professor Ola Bratt
University of Gothenburg
Department of Clinical Sciences
Division of Urology

POPULAR SCIENCE SUMMARY OF THE THESIS

Cancer in the large bowel or rectum is called colorectal cancer (CRC), and it is the second most common type of cancer in women and the third most common type of cancer in men in Sweden. Both CRC and precursors to CRC that are called adenomas, can bleed. Fecal Immunochemical Test (FIT) is a test that detects small amounts of blood in the stool. FIT could be used to screen for blood and select individuals with a higher risk of cancer that need to undergo a full camera investigation of the large bowel. The purpose of screening is to find cancer at an early stage when it is possible to cure the disease. Usually, FIT screening is repeated every second year. How well FIT performs depends on the population, the number of samples and the threshold chosen for a positive test. Previous evaluations have shown that FIT screening performs better in men than in women.

This thesis is about how to select CRC screening-individuals for further bowel investigation in the most beneficial way. How many FIT-samples should be taken? Which threshold should be used? Which cancers and adenomas are detected or missed with FIT? Should there be different screening-regimens for women and men?

In paper I we evaluated two FIT samples at different cut-off levels in a FIT-positive cohort from the randomized controlled study Screening of Swedish Colons (SCREESCO). We found that the FIT result was higher in individuals with CRC and advanced adenomas at bowel investigation, compared to those without findings. FIT was also higher in the presence of large adenomas compared to smaller ones. Moreover, one sample with a low threshold for positive test identified more CRC than two tests with a higher threshold. For the low thresholds, a positive test was better in predicting advanced adenomas in men than in women, but equally good at predicting CRC in men and women.

In paper II we investigated the sensitivity and specificity of FIT at different thresholds for a positive test and number of samples in a colonoscopy cohort from the SCREESCO study. Sensitivity is the probability that the test is positive when a person has the disease and specificity is the probability that the test is negative when a person is healthy. Sensitivity and specificity for CRC and advanced adenomas was 7-26% and 89-99% respectively depending on the number of samples and the chosen threshold. The lower the threshold and the more samples, the lower specificity and the higher sensitivity. FIT more often detected adenomas with high risk of becoming cancer and did so to a higher extent in men than in women.

In paper III we evaluated the screening strategy of the population-based Stockholm-Gotland screening program which since 2015 uses a lower threshold for a positive test in women as compared to men. With this strategy, significantly more CRCs were found in men than in women, and a normal bowel investigation was more common in women. However, almost 25% of the CRCs in the screened women would have been missed if they had been screened at the same threshold as men - with only minor savings of screening costs per detected CRC.

In paper IV we investigated the interval cancers (IC) of the population-based Stockholm-Gotland screening program. IC is a CRC detected between two screening rounds and missed

by the screening program. The incidence of IC in each age and gender group was compared to the CRC incidence (the new cases diagnosed) before initiation of the screening program. The test sensitivity for CRC was higher in women than in men but was estimated to be equal if the threshold for a positive test had been the same in both genders. The number of ICs per 10,000 negatively screened were significantly higher in men than in women with the Stockholm-Gotland screening strategy. However, when compared to the CRC incidence before screening implementation the differences in the ICs between genders were non-significant, suggesting that the CRCs are missed in the same rate in men and women as they are expected to appear in the population.

ABSTRACT

In Sweden, colorectal cancer (CRC) is the second and third most common type of cancer in men and women respectively. The relative five-year survival is approximately 65%, but prognosis is better if diagnosed at an early stage of disease. Fecal Immunochemical Test (FIT) detects blood in the stool and is used in screening, and individuals with a positive test are referred for colonoscopy. Several studies have indicated a lower sensitivity for advanced neoplasia (AN; CRC and advanced adenomas) in women as compared to men. In the Stockholm-Gotland region, population-based screening was initiated in 2008, and from 2015 FIT screening with lower cut-off levels for a positive test in women (40μg/g) than in men (80μg/g) was applied. The aim of this thesis was to increase the knowledge of the performance of FIT in an average-risk Swedish screening population and to explore a gender-specific screening strategy regarding colonoscopy findings, screening costs and interval CRC (IC; CRCs detected between two screening rounds after a negative screening episode).

In Paper I the performance of two FIT samples at different cut-off levels was evaluated in a FIT-positive cohort from the randomized controlled study Screening of Swedish Colons (SCREESCO). The FIT level was significantly higher in individuals with CRC and AA as compared to other participants and correlated to adenoma size. CRC detection increased with lower cut-off level and multiple samples and was significantly higher with one sample at a low cut-off level than two samples at a higher cut-off level. The positive predictive value (PPV) for AA was significantly higher in men than in women for one and two samples at cut-off levels <40μg/g but PPV for CRC was equal between genders at all cut-offs and number of samples.

In paper II the accuracy of two FIT samples at different cut-off levels were evaluated in a colonoscopy cohort from the SCREESCO study. Sensitivity and specificity for AN ranged from 7-26% and 89-99% respectively depending on the number of samples and the cut-off level. There was no gain in sensitivity using two samples instead of one, for any of the cut-off levels. Specificity was significantly higher with one sample as compared to any of the two samples, at the lowest cut-off levels. In the 225 participants with adenomas, pedunculated shape and high-risk dysplasia was independently associated with FIT positivity at cut-off ≥10μg/g for any of the two samples. Sensitivity for AA was significantly higher in men vs women, but specificity was similar between genders.

In paper III the Stockholm-Gotland population-based screening program was evaluated regarding colonoscopy findings and costs in a screening cohort from 2015-2017. CRC was found in significantly more men than women, 138 (8.3%) vs 120 (5.8%). A normal colonoscopy was more common in women than in men (24% vs 17%, p-value <0.05). Had the cut-off level been 80μg/g in both genders, the PPV for CRC was estimated to be equalized between genders. However, in women with CRC, 28 (23%) had FIT level of 40-79μg/g and would thus have remained undetected at cut-off level 80μg/g in both genders. The gender-specific screening strategy was estimated to be 16% more expensive than the gender-equal strategy, corresponding to a 3% increment in costs per detected CRC.

In paper IV the ICs were evaluated in the first round of the Stockholm-Gotland population-based screening program and compared to the experienced incidence rate (EIR) prior to screening implementation. In the cohort 124 FIT ICs, 7 colonoscopy ICs, 3 ICs in individuals non-compliant to colonoscopy and 177 CRCs in non-participants were detected within 2 years. Test sensitivity was 0.75 in women and 0.62 in men (p-value 0.011), but would have been equal, had cut-off level been 80µg/g in both genders. The IC rate was significantly higher in men than in women, 12.6 vs 6.0 per 10,000 negatives. The rate ratio of the IC incidence/EIR was 0.30-0.44 and non-significantly lower in the women as compared to the men in each age group. In all the 568 CRCs including those in non-participants, proximal localization was significantly more common in women (42%) than in men (29%).

LIST OF SCIENTIFIC PAPERS

I. **Fecal immunochemical test in colorectal cancer screening: Colonoscopy findings by different cut-off levels**
H Ribbing Wilén, J Blom, J Höijer, R Hultcrantz
Journal of Gastroenterology and Hepatology 2019;34:103-112

II. **Fecal immunochemical test in cancer screening – colonoscopy outcome in FIT positives and negatives**
H Ribbing Wilén, J Blom, J Höijer, G Andersson, C Löwbeer, R Hultcrantz
Scandinavian Journal of Gastroenterology 2019;54(3):303-310

III. **Gender-specific cut-off levels in colorectal cancer screening with fecal immunochemical test: A population-based study of colonoscopy findings and costs**
H Ribbing Wilén, D Saraste, J Blom
Journal of Medical Screening 2021;28(4):439-447

IV. **Interval cancers in a population-based screening program for colorectal cancer with gender-specific cut-off levels for fecal immunochemical test**
H Ribbing Wilén, D Saraste, J Blom
Journal of Medical Screening DOI:10.1177/09691413221085218 *online ahead of print*

CONTENTS

LIST OF ABBREVIATIONS

AA	Advanced adenoma
AN	Advanced neoplasia
ASA	Acetylsalicylic acid
BMI	Body Mass Index
CI	Confidence interval
CIMP	CpG island methylator phenotype
CIN	Chromosomal instability
CRC	Colorectal cancer
ESD	Endoscopic submucosal dissection
FAP	Familial adenomatous polyposis
FIT	Fecal immunochemical test
FIT IC	Interval CRC after negative FIT
FOBT	Fecal occult blood test
gFOBT	Guaiac fecal occult blood test
HNPCC	Hereditary nonpolyposis colorectal cancer
IC	Interval cancer
ICER	Incremental cost-effectiveness ratio
MMR	Mismatch repair
MSI	Microsatellite instability
MSS	Microsatellite stable
NNS	Number needed to screen; Number needed to scope
Non-AA	Non-advanced adenoma
NPV	Negative predictive value
NSAID	Non-steroidal anti-inflammatory drug
OR	Odds ratio
PPV	Positive predictive value
QALY	Quality adjusted life years
QoL	Quality of Life
RCC	Regional Cancer Center
SCRCR	Swedish Colorectal Cancer Register

SD-CRC	Screening detected CRC
SG	Standard gamble
TNM	Tumor Node Metastasis staging system
TTO	Time trade-off
VAS	Visual analogue scale
WEO	World Endoscopy Organization

1 INTRODUCTION

1.1 Colorectal cancer and adenoma

1.1.1 CRC epidemiology and prognosis

Worldwide, colorectal cancer (CRC) is the third most common type of cancer, diagnosed in approximately 1.9 million people every year and the cause of more than 900 000 deaths (1). In Sweden, CRC is diagnosed in approximately 6800 per year; the second most common type of cancer in women and the third most common in men. The lifetime risk of colonic CRC is approximately 2% at the age of 75. The median age at diagnosis for colonic CRC is 74 years and about 90% are over the age of 60 at diagnosis. The median age at diagnosis for rectal cancer is 71 years. The five-year relative survival rate for CRC has improved over the past decades to around 65%, however the prognosis is much better if detected at an early compared to late stage of disease; the relative survival for stage I colonic disease is approximately 95% and for stage IV disease around 15% (2)(3)(4).

1.1.2 Risk factors for CRC

Several risk factors have been identified in the development of CRC. Most CRCs are sporadic and only approximately 3-5% attributable to hereditary syndromes, i.e., Lynch syndrome (Hereditary non-polyposis colon cancer; HNPCC) and Familiar adenomatous polyposis (FAP), but 35% of the CRC risk is estimated to be due to hereditary, largely unknown factors. Inflammatory Bowel Disease (IBD) confers a higher risk of CRC, as do male sex and old age. Lifestyle factors such as high intake of alcohol and red meet, smoking, diabetes, and obesity are associated with a higher risk, and physical activity, high intake of fibers and vegetables is thought to be protective. Hormone replacement therapy and regular aspirin intake is associated with a 20-30% lower CRC risk (4).

1.1.3 CRC stage and treatment

CRCs are classified in the TNM (Tumor Node Metastasis) system, in which T represents the depth of invasion of the tumor, N is the extent of the spread to regional lymph nodes or presence of peri colorectal tumor deposits, and M the presence and extent of distant metastasis. Thus, a T1 CRC invades the submucosa, T2 the muscularis propria of the bowel wall, T3 the subserosa, and T4 grows into other organs or the visceral peritoneum, with further subclassifications. Stage I refers to T1-T2 CRC, stage II to T3-T4 CRC, stage III exhibit regional lymph node metastases and stage IV distant metastasis (5). The keystone of CRC treatment is surgery of the affected segment of the bowel, that could be performed either as open or minimally invasive, e.g. laparoscopic or robotic, resection (6). However, for superficial T1 CRC, endoscopic resection, e.g., endoscopic submucosal resection (ESD), is feasible. In colonic CRC stage III, and stage II with high risk of recurrence, neoadjuvant chemotherapy is advocated (7). For rectal cancer, preoperative radiotherapy is used to reduce the risk of local recurrence, and locally advanced rectal cancer is treated with preoperative chemoradiotherapy (8). In stage IV CRC surgery is possible if a complete removal of

metastasis e.g., in liver and lung, is achieved and in combination with chemotherapy, and when unresectable, different chemotherapy regimens depending on the tumor molecular profile can lengthen the survival (9).

1.1.4 The adenoma-carcinoma pathway

About 70-85% of CRCs are developed stepwise from normal mucosa to adenoma and carcinoma, via a process called the chromosomal instability- (CIN) or suppressor pathway. The non-advanced adenoma (non-AA) grows from an aberrant crypt focus, and gradually displays a higher rate of villous architecture and dysplasia. Dysplasia is graded in high grade (HGD) or low grade (LGD) according to the microscopic appearance. When a HGD lesion invades through the muscularis mucosa layer in the bowel wall, the lesion is defined as an adenocarcinoma (10). Some of the genetic events associated with the pathway are mutations in the proto-oncogene K-RAS, deletion or mutation of the APC-gene and the deletion of the tumor suppressor gene p53, which gives rise to chromosomal instability (CIN), a development that is estimated to take 5-15 years (11-13) (Fig. 1).

The mutator or microsatellite instability (MSI) pathway to CRC is seen in Lynch syndrome as well as in about 10-20% of sporadic CRC. This pathway involves mutations in mismatch repair (MMR) genes that give rise to microsatellite instability (MSI). Microsatellites are small repeated sequences of DNA, and genomic instability occurs with defect MMR genes leading to an inaccurate number of copies of these regions. MSI is graded in MSI high (MSI-H), MSI low (MSI-L) or microsatellite stable (MSS) depending on the number of defect regions that are detected (13, 14). MSI-H CRC are more often found in proximal colon and in women (15, 16).

A third pathway in CRC development is inappropriate methylation of the genome, for example CpG islands (Cytosine-Guanine dinucleotide group) promotor regions, which can lead to a dysfunction of tumor suppressor genes, classified as CpG Island Methylator Phenotype (CIMP) found in about 20-30% of CRCs (17, 18).

Serrated polyps include hyperplastic polyps (HP), traditional serrated adenoma (TSA) and sessile serrated lesions (SSL) and could be difficult to detect at colonoscopy due to their flat appearance, mucus-covered surface, and proximal location in colon. SSL and TSA are considered precursors to 20-25% of sporadic CRC and features mutations in the K-RAS or proto-onco gene BRAF, CIMP, and sometimes MSI, and could present a faster turn-over from precursor lesion to CRC (13, 18, 19) (Fig.1).
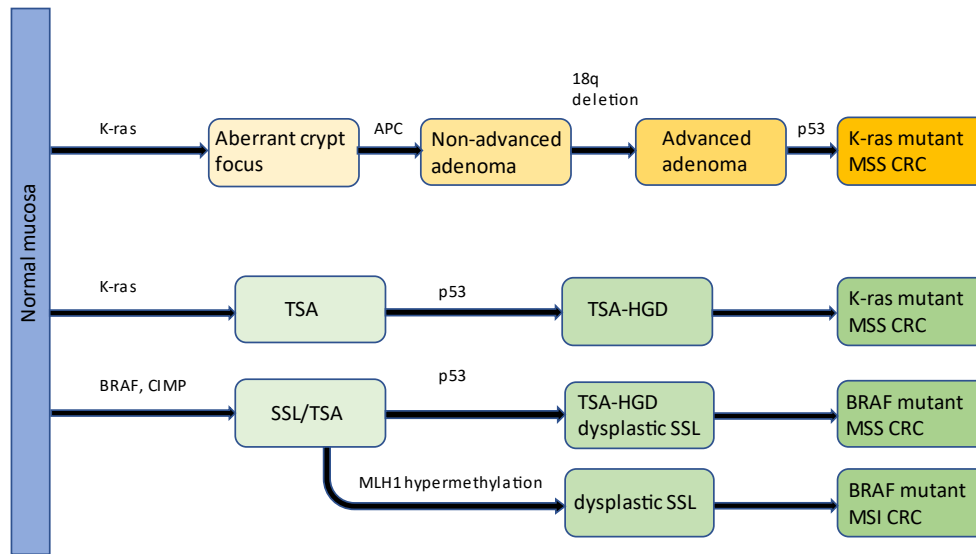
Figure 1. Simplified schematic illustration of the CIN and serrated pathway to CRC development. (Adapted from Worthley D.L. et al World J Gastroenterol 2007;13(28):3784-91 and Crocket S. et al Gastroenterology 2019;157(4):949-966)

1.1.5 Advanced adenomas and serrated polyp surveillance

Advanced adenomas (AA) are defined as adenomas ≥10mm, ≥3 adenomas or adenomas with villous histology or high-grade dysplasia (HGD) and confers a higher risk of developing CRC (20). In a pooled analysis of 9 000 American patients, those with ≥3 adenomas had twice the risk, and those with adenomas ≥10mm a two- to three-fold increased risk of CRC or AA at follow-up colonoscopy, and increased risk was also seen for proximal location and villous histology (21). However, most adenomas do not progress to CRC: the cumulative incidence of CRC in patients with polyps ≥10mm was 2,5%, 8% and 24% after 5, 10 and 20 years of surveillance according to an observational study conducted before the introduction of colonoscopy and polypectomy (22). European guidelines recommend that patients with adenomas 10-20mm or ≥3 adenomas (and possibly those with adenomas featuring HGD and villous histology) undergo a control colonoscopy after 3 years and those with adenomas ≥20mm or ≥5 adenomas after one year (23).

In an American study of 6 700 screening colonoscopies performed by university hospital gastroenterologists, 13% exhibited sessile polyps in the proximal colon, but the detection rate varied greatly between endoscopists (24). In European colonoscopy screening cohorts, the detection of sessile polyps was between 9-12% (25). According to the Swedish polyp surveillance program, sessile polyps ≥10mm and sessile polyps with dysplasia regardless of size are regarded as high-risk adenomas requiring a follow-up colonoscopy after 3 years (26). European guidelines recommend sessile polyp surveillance as for other adenomas (23).

1.2 CRC screening

1.2.1 FIT and gFOBT

Both CRC and AA may bleed. Fecal Occult Blood Test (gFOBT, e.g., Hemocult©) is a guaiac based qualitative test for detecting blood in stool. The heme in the Hemoglobin oxidizes guaiac and the chemical reaction is read visually. Dietary components could give a false positive test, for instance intake of red meet and Vitamin C could give a false negative result. Fecal Immunochemical Test (FIT) is an immunoassay with antibodies directed against the globin part of the Hemoglobin. The antibody-antigen complex is analyzed with turbidimetry, and thus gives a quantitative measure of the amount of blood in the sample. The antibodies are specifically directed to human globin, hence dietary restrictions are not necessary, and as globin is degraded in the upper gastrointestinal tract the assay detects colorectal bleedings. There are also qualitative FITs that read positive at a specific cut-off value (27-29). Several kinds of quantitative FITs exist but test performance may vary between brands even if the same cut-off level is used (30).

A meta-analysis of 19 studies of FIT accuracy for detecting CRC in asymptomatic adults showed a pooled sensitivity of 0.79 and specificity of 0.94, but the results varied with the chosen cut-off level for a positive test. A low threshold and to a lesser extent multiple tests rendered a higher sensitivity and lower specificity. In 12 of the 19 included studies, colonoscopy was used as the reference for detecting CRC (31). In a more recent meta-analysis of 31 studies, CRC sensitivity ranged from 0.71-0.91 and specificity from 0.90-0.95 depending on the cut-off level (32).

Several studies have compared the accuracy of gFOBT with that of different brands of FITs for the diagnosis of advanced neoplasia (AN; advanced adenoma + CRC) using colonoscopy as reference. If applying a FIT cut-off that renders the same specificity as that of gFOBT, the FIT sensitivity for AN was between 26% (1 sample, cut-off 20µg/g) to 44% (3 samples, cut off 20µg/g), as compared to 9-20% for gFOBT. At a specificity of 92% and 95% the FIT sensitivity for CRC was 73% (1 sample cut-off 20µg/g) and 85% (2 samples, cut-off 20µg/g) as compared to 33% and 31% for gFOBT (33-35). Other studies have compared qualitative or semi-qualitative FIT to gFOBT (36).

1.2.2 Endoscopy

Endoscopy is a camera examination of the intestine; colonoscopy includes the entire large bowel from terminal ileum to rectum, and sigmoidoscopy is limited to the descending/sigmoid colon to rectum (Figure 2). Colonoscopy is regarded as the golden standard of large bowel assessment; a positive screening FOBT test is followed by a colonoscopy to investigate the source of the bleeding, and a positive screening sigmoidoscopy is followed by a complete colonoscopy to detect synchronous large bowel lesions, but colonoscopy could also be used as the primary screening test. For a complete inspection of the mucosa, a clean bowel is required. There are several regimens of laxatives to prepare for colonoscopy and a split dose is recommended: one in the evening before and

one in the morning on the day of the examination. Regarding sigmoidoscopy a single dose of enema is sufficient (37).

Serious complications to colonoscopy could occur from the barotrauma caused by the pressure to the bowel wall and from procedures of polypectomy, and includes bleeding, perforation and even mortality. The mortality could be related to complications to the colonoscopy itself but in most cases due to comorbidity such as cardiovascular disease and cirrhosis (38).
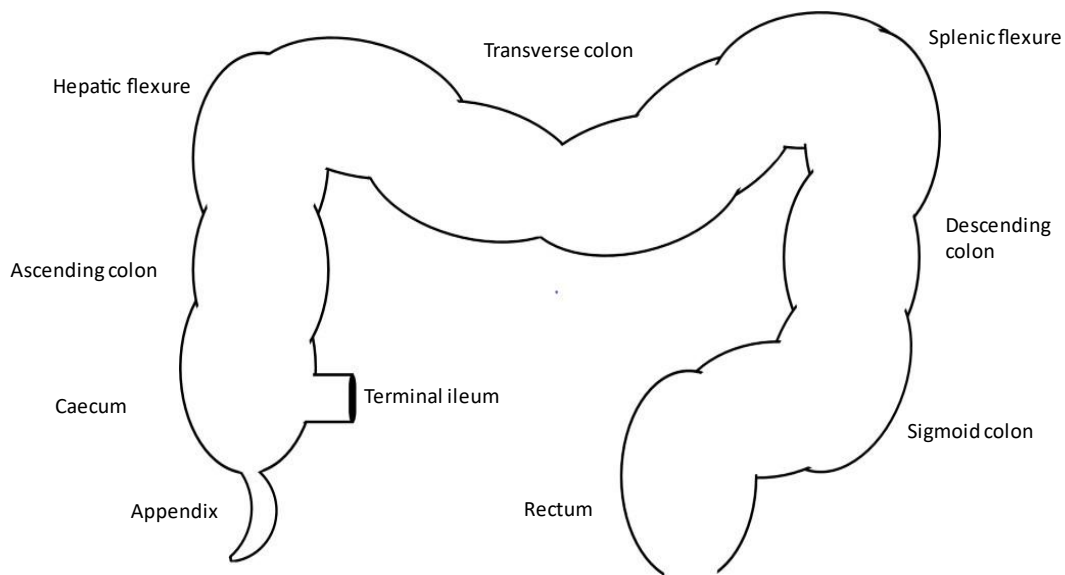
Figure 2. Schematic illustration of large bowel anatomy.

1.2.3 Other screening methods

CT colonoscopy is a radiological large bowel examination, a "virtual colonoscopy", that most often requires a full bowel preparation and/or fecal tagging in combination with bowel insufflation to visualize the mucosa. Those with significant lesions (usually ≥6-10mm) are further investigated with colonoscopy. The detection of AN ≥10 mm is comparable to that of colonoscopy, but lower for flat lesions. Moreover, the possibility of diagnosing extracolonic lesions and a low complication rate constitutes the advantages of the method. Potential disadvantages include the risk of radiation-induced cancer in the screened population, and the demand for follow-up colonoscopy - preferably in the same day to avoid the need for an extra bowel preparation (39). Randomized trials have demonstrated a higher participation rate as compared to colonoscopy screening (40).

Magnetic Resonance Imaging (MRI) is another non-invasive imaging screening option that avoids ionizing radiation but is contraindicated for those with metallic implants and often implies longer examination time and suffer from limited availability (41).

Evaluation of Multi-target stool DNA test that combines FIT with detection of mutated K-RAS and inappropriately methylated BMP3 and NDRG4 regions have showed an increased sensitivity but an inferior specificity for CRC and AN as compared to FIT alone (42, 43). Due to the high costs, almost all other screening strategies were more cost-efficient in a simulation model (44).

1.2.4 Randomized controlled studies on screening with gFOBT and endoscopy

The aim of CRC screening is to reduce disease-specific mortality. To reach this aim with screening the disease needs to be common and curable at early detectable stages, and the screening method to be accepted in the target population (45). These criteria are possible to fulfil in CRC screening, and screening is also recommended for breast and cervical cancer by the EU (46). Overall mortality is not expected to be reduced by screening, because CRC accounts for only about 3% of all deaths in Sweden (47).

There are four randomized controlled studies that have compared screening with gFOBT to no screening; the studies by Mandel et al (Minnesota trial), Kronborg et al (Funen, Denmark), Hardcastle et al (Nottingham trial) and Lindholm et al (Goteborg)(48-51). The study protocols differed in several aspects between these trials, e.g. in study population (volunteers in the Minnesota trial and invitational in the Goteborg trial, 60-64 year-olds in Goteborg and wider age range in other trials), gFOBT testing (rehydrated gFOBT in the Minnesota and Goteborg trials, re-testing both positives and negatives in the Nottingham trial), screening interval (annual and biennial in the Minnesota trial, cohort with 10 years re-screening interval in the Goteborg trial). A meta-analysis showed a 16% lower cumulative CRC mortality in the screened populations (52).

Screening with sigmoidoscopy compared to no screening has been evaluated in five RCTs; Atkin et al (UK), Segnan et al (SCORE, Italy), Schoen et al (US), Thiis-Evensen et al (Telemark, Norway), Holme et al (NORCCAP Telemark and Oslo, Norway), in which attendants with (advanced) findings at sigmoidoscopy were referred for total colonoscopy (53-57). Participation rate was between 58-83%, but it is noteworthy that in the Italian and UK study only individuals reporting willingness to attend screening were included in the randomization. The Italian, NORCCAP and Telemark study failed to prove a significant reduction in disease specific mortality at early follow-up between 7-11 years, but a meta-analysis showed a reduced CRC mortality of 28% (58, 59).

Colonoscopy screening has not yet been evaluated for CRC mortality in randomized controlled studies, but the ongoing NordICC study on colonoscopy vs no screening has reported a lower uptake (40%) compared to FIT and sigmoidoscopy (60). The ongoing COLONPREV, CONFIRM and SCREESCO are randomized studies that compares FIT

screening to colonoscopy screening with regards to disease mortality (61, 62)(63). In addition, there is also an ongoing Norwegian randomized trial comparing repeated FIT screening to sigmoidoscopy (64).

1.2.5 Established screening-programs and recommendations on screening strategy

US Preventive Services Task Force as well as American Cancer Society recommends CRC screening with FOBT or endoscopy starting at the age of 45 (65, 66). In the European guidelines screening is recommended in 50-74-year-olds with gFOBT, FIT or sigmoidoscopy in order to decrease the disease mortality (46). By the year 2008, 12 out of 22 European countries had established regional or national screening programs, and by 2015 this had increased to 24 of 28 countries (67, 68).

1.3 Screening measures and bias

1.3.1 Screening detected CRCs and overdiagnosis

When a screening program with repeated screening rounds is implemented, it is expected that the rate of newly diagnosed CRC, i.e., the CRC incidence, increases as both the new cases and the cases already present in the population, i.e., the prevalent CRCs, are diagnosed in the first round. During the subsequent screening rounds the CRC incidence should return to the previous background incidence or lower if CRC precursor lesions are removed (69).

Overdiagnosis is a general concern in screening and refers to an asymptomatic diagnosis by screening that would otherwise not have caused morbidity or mortality within the lifespan of an individual. Overdiagnosis is harmful by causing anxiety and leading to unnecessary surveillance and overtreatment. FIT screening primarily aims at detecting early stages of CRC and thereby improving disease mortality. Screening directed to detect and remove adenomas could reduce the CRC incidence, but at the expense of overdiagnosis and overtreatment since most adenomas do not progress to CRC (70). Overdiagnosis is difficult to quantify but could be estimated as the difference between the cumulative incidence in a screened population as compared to an unscreened. Although CRC incidence increase with age, overdiagnosis and overtreatment is of particular interest in an older population because of a large burden of comorbidities and competing causes of deaths other than CRC (71). An attempt to quantify the overdiagnosis in gFOBT screening was made with five microsimulation models on long-term follow-up data from the Nottingham trial and was estimated to be between 0-7.6% (72).

1.3.2 Interval cancers in FIT screening

According to the World Endoscopy Organization (WEO) a screening detected CRC (SD-CRC) is defined as a CRC diagnosed after a positive screening episode, i.e., after a positive FIT and positive screening colonoscopy or after a positive screening colonoscopy if this is the primary screening test, and usually within 6 months from the screening. A non-screening detected CRC is defined as a CRC diagnosed either in a non-participant or in a participant

after a negative screening episode and before the next screening round, i.e., an interval CRC (IC). ICs are further classified as IC after a false negative FIT (FIT IC) or after a positive FIT and false negative screening colonoscopy (colonoscopy IC). ICs could also occur in FIT-positives that are not compliant with the screening colonoscopy. CRCs diagnosed in FIT non-participants are not defined as ICs (73). There is also a possibility that fast-growing de nuovo CRCs are diagnosed between the screening intervals, and that the FIT at the time of screening was truly negative (74).

The rate of ICs is dependent on the screening method used and the screening interval, but also on the CRC incidence of the background population, and the CRC incidence of the individuals willing to undergo screening. Therefore, it is recommended that ICs are reported as number of IC per 100,000 person-years of follow-up and as a proportion of the background incidence, i.e., as a proportional incidence (69).

1.3.3 Accuracy measures

The accuracy of a test to correctly identify those with the disease and without the disease is expressed in the test sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV).

The test sensitivity is the proportion of test positive among those with the disease, TP/(TP+FN) in Table 1. A high test sensitivity means that a large proportion of the sick are identified by the test but does not consider the number of false positives. The test specificity is calculated as the number of test negatives among all healthy subjects, TN/(FP+TN) in Table 1. A high test specificity means that a large part of the healthy subjects is correctly identified as such, but it does not account for the number of false negatives, i.e., the sick subjects with a negative test. In the example of a FIT screening study the FIT result is verified against the findings at colonoscopy, although lesions could be missed at colonoscopy.

	Test positive	Test negative
Sick	True positive (TP)	False negative (FN)
Healthy	False positive (FP)	True negative (TN)

Table 1. Outcomes of a binary test in healthy and diseased.

However, in a screening situation only the test positives are referred for colonoscopy, and the number of sick are not known – and this is also the case in most clinical situations. The PPV describes the proportion of true positives, i.e., individuals with the disease, among all test positives, TP/(TP+FP) in Table 1. A high PPV indicates that most of the test positives truly have the disease. Conversely, the NPV is the number of healthy among all test negatives, TN/(TN+FN). A high NPV ensures that the test negatives are truly healthy. However, the predictive values are dependent on the prevalence of the disease in the population.

Disregarding the test sensitivity and specificity, the probability of being false positive is lower in a population where disease is common as compared to in a population where the disease is rare and vice versa, illustrated in Figure 3. (75)(76).
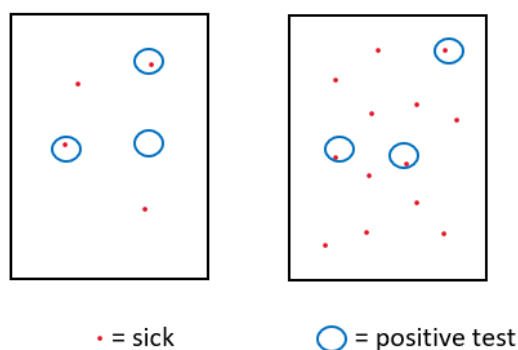
Figure 3. Disease prevalence in two different populations.

1.3.4 Bias in screening

There are several situations where bias can be introduced in screening, which is important to consider when evaluating results from screening studies. Firstly, self-selection bias occurs when a group of people with different outcome risk than the general population is more likely to participate in the study, thus the choice to participate in a study is related to the outcome and the risk of the outcome would be over- or underestimated in the study (77). As an example, results from randomized controlled studies are not necessarily reproducible in general screening settings. The study population may be different from the target population in the sense that volunteers in a screening study might be healthier than the general population, the "healthy volunteer effect", hence it is essential to evaluate ongoing screening programs in observational studies (78). Furthermore, there could also be a selection of individuals willing to attend screening whose CRC risk is lower than that of non-participants, thereby overestimating the effect of attending screening on disease outcome (79).

Secondly, concerning disease stage and survival the lead time bias must be considered. Screening aims at detecting CRC at an early stage, hence the time until disease progression or death is longer for screening detected CRCs than for CRCs diagnosed clinically at a later stage, but the total survival time for each stage might be the same. However, the goal of screening is to reduce the disease mortality which could be achieved by shifting a proportion of the clinically diagnosed late-stage CRC to a curable early-stage CRC.

Length-biased sampling is a special case of time bias, because slowly growing and hence less aggressive cancers are more likely to be captured in screening than clinically diagnosed, obscuring the survival rates when comparing the two (80).

1.3.5 Efficacy and effectiveness

The efficacy of screening is ultimately measured as a reduction in disease mortality, i.e., the risk difference or relative risk of CRC death between those invited and not invited to

screening seen in randomized controlled studies under optimal conditions. The effectiveness reflects the magnitude of the efficacy in an observational study, e.g., a routine screening program, also taking non-adherence and normal healthcare settings into account (81). Effectiveness could be expressed as the Number Needed to Screen (NNS) i.e., the number needed to be invited, to prevent one death or to detect one CRC. NNS is the inverse of the risk difference, or the inverse of the CRCs detection rate. The effectiveness depends partly on the accuracy of the screening test, but most importantly on the participation in and compliance with the screening program, since only the invited that participates in screening could conduce to the decreased mortality (82).

1.3.6 QALY and cost-efficiency

The advantage of an early-stage CRC diagnosis detected at screening is not only due to the ability to decrease disease specific mortality, i.e., the life years gained in a population offered screening, but also to a perceived gain in quality of life in patients diagnosed earlier as compared later with a disease at a more severe disease stage. Hence, Quality Adjusted Life Years (QALY) also takes the quality of life (QoL) into consideration and is used in health-economic evaluations. A health state that impairs the QoL is assigned a utility weight <1 and multiplied with the estimated survival time, which enables comparisons of health effects for e.g., different treatments, methods, or screening modalities. The utility weights assigned to the different disease stages could be estimated directly or indirectly, for hypothetical or self-experienced disease outcomes. Direct methods include Time Trade-Off (TTO), Visual Analogue Scale (VAS) or the Standard Gamble (SG) method. In SG the research subject compares the disease stage (e.g. CRC stage III with a stoma) to a range of probabilities of perfect health vs immediate death (e.g. 40% chance of perfect health and 60% risk of immediate death), and when the two scenarios of the disease stage and the probability of perfect health are perceived equally attractive, this probability constitutes the utility weight (83). In the TTO method the research subject compares a certain amount of time spent in the disease stage until death with another amount of time spent in a perfect health until death, and when these alternatives are viewed as equally preferable, the ratio of the time in perfect health divided by the time in disease stage constitutes the utility weight (84). There are also indirect methods of determining the utility weight in which the results of QoL questionnaires are translated to utility weights. In a large Swedish study the self-experienced TTO and VAS was validated against a QoL questionnaire, and the mental health status was the dimension mostly influencing the TTO and VAS (85).

The Incremental Cost-efficiency Ratio (ICER) is the cost per gain in QALY comparing one treatment or method to another (86). Willingness-to-pay is the threshold regarded as cost-efficient and varies between countries. In Sweden the National Board of Health and Welfare recommend a cost of 100,000- 500,000 SEK (≈8,100-40,300 £) per QALY gained (87). The National Institute of Health and Care Excellence (NICE) in the United Kingdom has proposed a threshold of 20-30,000£ per QALY gained (88). According to WHO a threshold of ≤1-3 times the BNP per capita per QALY gained is viewed as cost-efficient (89, 90).

In Markov Models each health stage is associated with a certain utility weight and cost, and the risk of progression from one stage to another is calculated, e.g., the risk of AA to progress to CRC over a certain time and all costs associated with the change. Costs and QALYs gained are then summarized from all the possible outcomes and compared between the different treatment strategies (91).

2 LITERATURE REVIEW

2.1 FIT performance

2.1.1 Sensitivity and specificity for AN and AA

Several studies have evaluated the accuracy of FIT for AA and AN, but the study populations, the types and brands of the FIT tests, the screening strategy and sometimes the definitions of AA and AN differ between studies. Sensitivity and specificity for AA depends on the study population, the number of samples and the cut-off level for a positive test, and the adenoma characteristics. The participation rate in screening studies and screening programs differ between countries, and the definition of participation varies between studies (82). The screening settings in different countries also vary regarding the level of care and endoscopist and pathologist (sub-) specialization (92). It is therefore difficult to pool the results of FIT accuracy for AA into one estimate and meta-analyses suffers from heterogeneity (32). Short summaries of studies of relevance for this thesis are listed below.

Colonoscopy screening is established in the south of Germany. In the BLiTz study cited previously, screening participants were recruited for evaluation of different FIT and gFOBT brands at 20 gastroenterology units 2005-2009. Brenner et al compared FIT accuracy from 2,200 participants 50-79 years of age. One fecal sample was assessed with RIDASCREEN® hemoglobin (cut off 24.5µg), RIDASCREEN® hemo-/haptoglobin complex (cut off 7.95 µg) and OC Sensor (cut off 6.1µg), which rendered a positivity rate of 5%. The FIT cut-off levels were chosen to give the same positivity rate as the gFOBT. FIT sensitivity and specificity for AN was 20.3-25,7 and 96.8-97.4 respectively, and that of all neoplasias (CRC and all adenomas) 10.2-12.1 and 97.1-97.8 respectively (34).

In Hong Kong, Wong et al compared one and two samples of qualitative FIT (Hemosure®, cut off 10 µg) in 50-70-year-olds invited to FIT-based screening at two tertiary hospitals. Each participant returned two fecal samples and one of these was randomly selected as the one-sample strategy. In 5,300 participants, the positivity rate was 7.2% for one sample and 7.8% for two samples. Sensitivity and specificity for AA was 33.1 and 91.5, and that of AN was 34.7 and 91.7, and results did not significantly differ between one- and two-sample strategy (93).

A retrospective analysis was conducted in 21,800 participants of a health program in two hospitals in Kameda, Japan 1983-2002. The health program included colonoscopy and FIT testing with one qualitative FIT (Magstream 1000/Hem SP, cut-off 100-200 µg/g). The participants were predominantly young (60% <50 years) and male (72%). The positivity rate was 5.6% and sensitivity and specificity for AN was 27.1 and 95.1 respectively (94).

FIT accuracy was assessed in a prospective Korean study of 3,800 asymptomatic individuals aged 15-78 and 300 CRC-patients referred to the National Cancer Center. A qualitative FIT (OC Hemodia, cut-off 20 µg) was used, and 1.4% in this study population was positive. Sensitivity for AA was 6% but specificity for AA was not reported (95).

At Taiwan University Hospital asymptomatic individuals underwent colonoscopy as part of a health check-up. Of these, 18,200 participants >50 years provided one FIT (OC Light, cut off 10 µg) of which 7.3% were positive. Sensitivity for AA was 28% and specificity 93.5% (96). At another Taiwanese hospital (Far Eastern Memorial Hospital), 2,800 volunteers, 19-84 years old, in a health program underwent both gastroscopy, colonoscopy and FIT (OC-light, cut off 10µg). In this group, 14% were FIT positive, and sensitivity and specificity for adenoma was 21.4 and 88.9 and that of all neoplasia (CRC and adenoma) 24.8 and 88.9 respectively. Lesions in the upper gastrointestinal tract was not associated with positive FIT (97).

Siripongpreeda et al recruited 1,400 50-65-year-olds in Bangkok for colonoscopy investigation and concomitant FIT (FOB one-step test, Abon Biopharm, limit of detection 6 µg). In this cohort, 69% were women, 4% were symptomatic, and 8% had heredity for CRC. With this selection CRC was detected in 1.3%, and sensitivity and specificity for CRC was 56% and 96% respectively (98).

The COCOS trial was conducted at two endoscopy units in the regions of Amsterdam and Rotterdam to compare screening with CT colonography to primary colonoscopy. A cohort of 1,260 50-75-year-olds in the colonoscopy arm also provided FIT (OC Sensor, cut-off analyzed at 10, 15 and 20µg). Sensitivity and specificity for AA was 29-35% and 97-93% respectively depending on cut-off level (99).

Yuan et al evaluated FIT performance in 700 45-75-year-old participants of a health program at three medical centers in China who underwent colonoscopy and FIT testing (OC Sensor, Eiken, Japan, cut off 30 µg). The sensitivity for AN was considerably higher at 65% with a specificity of 32%. However, 41 (6%) CRCs were detected in the cohort, which is much higher than in other screening cohorts, due to a selection of those with high FIT-levels willing to undergo colonoscopy (100).

2.1.2 Single or multiple samples?

Park et al evaluated FIT accuracy in a cohort of 770 50-75-year-olds participating in colonoscopy screening at four tertiary hospitals in South Korea. The participants provided three consecutive FIT samples (OC Sensor, cut-off 10-30µg/g), and three gFOBT. For a single sample at cut-off 20µg the sensitivity and specificity for AA was 24% and 94% respectively. With two and three samples at cut-off 20µg the sensitivities and specificities were 28% and 92% and 35% and 90% respectively (35).

Rozen et al and Levi et al investigated the accuracy of FIT in two partly overlapping cohorts with 1,200 symptomatic low-risk or asymptomatic high-risk patients referred for colonoscopy. They also provided three FIT samples (OC Sensor, cut-off 10-30 µg/g) yielding a positivity rate of 9-13% depending on the number of samples. The sensitivity and specificity for AA for one fecal sample was 23-38% and 97-93% respectively depending on cut-off level. At cut-off 20 µg/g the sensitivity and specificity with two and three fecal samples was 38 and 44% and 95% and 93% respectively (101, 102).

A randomized controlled study from the Netherlands compared screening with one vs two FIT samples (OC Sensor cut-off 10 µg/g). Participation was equal in both study arms (61%), and at a positivity rate of 3,2-6,2% the detection rate of AN was also equal between the one- and two sample strategies. The detection rate of AA increased proportionally more than that of CRC when adding a second sample, implying that AA bleeds more intermittently than CRC (103). When the same study cohort was invited to repeated screening the cumulative detection rate of AN was the same in both study arms, hence the single sample strategy was advocated (104, 105).

In a French study gFOBT was compared to FIT in screening participants aged 50-74 years. Two fecal samples (Magstream, cut-off 100-200µg) was used, which rendered a positivity rate of 8% and 1,300 colonoscopies. For the single-sample strategy a randomly selected sample of the two was used, and the mean of the log-transformed Hemoglobin level of the two samples was assessed. Those with a negative test were not investigated with colonoscopy, hence the ratio of the true positives and false positives between the FIT the gFOBT was calculated instead of sensitivity and specificity. The mean of the two samples rendered the largest increase in true positive ratio. The cut-off for FIT was chosen as to give the same positivity or specificity as that of gFOBT (i.e. same number of false positives) (106).

In four regional screening programs in Italy, participants 50-69 years old were offered two fecal samples (OC Hemodia, cut-off 16, 20 and 24µg/g). The second test of the two was used as the single-sample strategy. The positivity rate was 2-8% depending on screening strategy, rendering 1,400 colonoscopies. At cut-off 20µg the detection rate of AA increased with 26% with two fecal samples as compared to one, and the corresponding increase for CRC was 21%. At cut-off 16 µg/g the detection rate of AA increased with 26% with two fecal samples as compared to one, and the corresponding increase for CRC was 17% (107).

In the colonoscopy arm of COLONPREV study performed in three tertiary hospitals in Spain, 779 participants 50-69 years old also provided two FIT samples (OC Sensor, cut-off 10-40µg). The first fecal sample and the highest of the two were assessed at different cut-off levels, which rendered a positivity rate of 5,8-13%. The sensitivity and specificity for AN was 28-35% and 95-97% for one sample strategy depending on cut-off level, and 28-42% and 91-95% for two samples respectively (108).

2.1.3 FIT and adenoma characteristics

Several studies have addressed how FIT performance is influenced by adenoma characteristics, i.e., size, localization in colon, number of adenomas, histology, and morphology.

Adenoma size

Park et al demonstrated a significantly higher FIT level in those with adenomas ≥10mm (35). Likewise, Rozen et al in the previously cited study of symptomatic low-risk or asymptomatic

high-risk patients referred for colonoscopy found a higher FIT level in those with large adenomas. Moreover, villous histology, HGD, distal localization and pedunculated shape was strongly correlated to adenoma size (101). In the young and mainly male colonoscopy cohort, Morikawa et al found a higher FIT sensitivity for large (>9mm) as compared to small adenomas (22% vs 7%) (109).

Adenoma localization

In the cohort of Park et al, participants with proximal AAs displayed a higher FIT level than those with distal AAs, however the proximal lesions (including the CRCs) were larger than the distal ones and there was no definition stated of proximal and distal localization (35). The above cited study by Chiu et al demonstrated a lower FIT sensitivity for proximal (22.5%) as compared to distal (31.6%) AAs, even after stratifying for flat and polypoid morphology (96). Likewise, in the cohort of Morikawa et al there was a lower FIT sensitivity for the large proximal adenomas than for the distal (11.2% vs 24.5%), but no difference in sensitivity for localization of the small adenomas (94, 109). Wong et al in the study from 2015 demonstrated a test sensitivity of 25% and 40% for proximal and distal AA respectively, and the differences remained regardless of using one or two FIT samples (93).

On the other hand, Wijkerslooth et al found equal sensitivity for proximal as for distal AN in the COCOS trial (99). Moreover, in the cohort of Rozen et al there was no difference in FIT levels between participants with proximal and distal AAs (101).

Haug et al evaluated FIT sensitivity for proximal and distal AN in participants of the BLiTz study and demonstrated a higher FIT sensitivity for distal than for proximal localization at all cut-off levels, possibly due to the higher rate of pedunculated shape among the distal lesions. When the analysis was restricted to those with a single AN, the difference in sensitivity remained only at the low cut-off levels (at <90% specificity). The result of this restriction was interpreted as a difference in test sensitivity between proximal and distal weaker sources of bleeding, but for profuse bleedings the localization did not affect the sensitivity (110).

Number of adenomas

Some of the aforementioned studies showed a correlation between FIT level and number of adenomas; Park et al found significantly higher FIT in those with ≥3 adenomas as compared to <3 adenomas, and Rozen et al demonstrated that number of adenomas was independently associated to FIT level (35, 101).

Histology and morphology

In the study by Park et al, participants with HGD adenomas displayed higher FIT levels as compared to those with LGD, but there were only two participants with HGD (35). Rozen et al found no independent association with grade of dysplasia and FIT level, neither did Digby et al in a Scottish FIT screening cohort (OC Sensor Eiken, cut-off 80μg) nor Ciatto et al in the

Florence screening program (OC-Hemodia Eiken, cut-off 20µg) in which the FIT-levels were associated with adenoma size (101, 111, 112).

As to the morphology, the FIT sensitivity seems higher for pedunculated or polypoid adenomas as compared to flat or broad based (96, 110). In a cohort of participants with adenoma from the COLONPREV study, FIT positivity (OC Sensor, cut-off 10 and 20µg) in the first of the two samples was independently associated with pedunculated shape at both cut-off levels (113).

Participants with SSLs were evaluated in a Dutch study, and displayed similar FIT levels as those with non-AA or normal colonoscopy, and were equally common (10-15%) in colonoscopy screening cohorts as in those screened with FIT, because these lesions are less prone to bleeding (114). Chang et al evaluated colonoscopy screening in Taiwan and likewise concluded that the FIT result was similar in those with non-AA, normal colonoscopy, or SSLs (115). However, it is worth noticing that in the Dutch study only the most severe lesion was analyzed, hence synchronous adenomas could have contributed to the bleeding. Secondly, comparisons of colonoscopy and FIT screening cohorts should be performed with caution since the participation rates differed significantly; 22% in the colonoscopy cohort as compared to 52-57% in the FIT cohorts.

2.1.4 FIT and CRC characteristics

In many screening studies the number of CRCs detected are small which precludes detailed evaluations of CRC characteristics and FIT performance. Several meta-analyses have been performed regarding proximal vs distally located CRCs (and AN) with conflicting results; and in these meta-analyses studies using different FIT brands and cut-off levels are pooled together.

Haug et al included five FIT studies with colonoscopy as golden standard that reported on 71 CRCs altogether, and no pooled estimates of site-specific CRC sensitivity could be calculated (116). Hirai et al included 11 FIT studies on both symptomatic and average-risk individuals that underwent colonoscopy, and reported a significantly lower pooled sensitivity for proximal vs distal CRC of 71% and 79% respectively (117). However, a more recent meta-analysis by Lu et al evaluated 29 FIT studies, most of which were conducted in a screening setting, and found equal sensitivity for proximal as for distal CRC (67% vs 72%) (118).

Niedermaier et al conducted a meta-analysis of FIT sensitivity according to CRC stage including 44 studies that covered both screening, case-control, and symptomatic cohorts with colonoscopy as golden standard. A higher rate of early-stage CRC (stage I&II) was seen in the 12 screening studies as compared to case-control and symptomatic cohorts (70% vs 54% and 57%). The pooled sensitivity for stage I, II, III and IV CRC was 73%, 80%, 82% and 79% respectively - significantly lower in stage I vs the other stages. Moreover, the 9 studies that reported on T-stage showed substantially lower sensitivity for stage T1 than for T2, T3 and T4 (40% vs 79%, 83% and 66%). The occurrence of distant metastases or regional lymph node metastases determinant for stage III and IV CRC is likely not as related to the degree of

intestinal bleeding as the size of the primary tumor reflected in the T stage. The authors further hypothesize that the lower estimate for stage IV (and T4) vs stage III CRC might be related to general anemia in advanced stage cancer (119).

2.2 Gender differences and tailored screening

2.2.1 Gender difference in AN prevalence

In a meta-analysis of 18 colonoscopy screening studies with approximately 900,000 individuals >40 years of age, the detection rate of AN was 2,6-17% in men and 1,8-9,3% in women. The prevalence differences corresponded to an 83% higher probability of detecting AN at screening colonoscopy in men than in women and was consistent in all age groups. The probability of detecting CRC was twice as large in men as in women. The high prevalence in men as compared to women could be related to hormonal and genetic factors or to differences in lifestyle between genders that are associated with the risk of developing CRC and AN such as smoking, alcohol use and obesity (120).

A more recent study from Austria of 44,300 screening colonoscopies (also including younger individuals with heredity for CRC) revealed a prevalence of adenoma of 20% (15% in women and 25% in men), AA of 6% (4,7% in women and 8% in men) and CRC of 1,1% (0,7% in women and 1,5% in men). The gender differences were significant and present in all age groups. A gender-equal prevalence was reached when the women were 10 years older than the men, which indicates that screening could be initiated at different ages in men and women (121). Brenner et al came to the same conclusion with an analysis of the cumulative incidence and mortality of CRC in different age and gender groups, as the same levels were reached 4-8 years later in life in women as compared to men (122).

The distribution of proximal and distal AN in men and women has been assessed in several studies, proximal location usually defined as lesions from caecum up to or including the splenic flexure. A comparison between a Veteran Affairs study (only males) and a female screening colonoscopy cohort revealed that the proportion of only proximal AN, i.e. without synchronous distal neoplasia, was larger in women than in men (2/3 as compared to 1/3) and therefore would be missed at sigmoidoscopy screening (123). On the contrary, in the COCOS trial the distribution of proximal and distal AN was equal in men and women, disregarded other concomitant neoplasia (124). For CRC, a study of 17 000 patients in Germany with colonic CRC demonstrated a higher proportion of proximal CRCs in women, and in this study the right-sided cancers also displayed a worse prognosis (125).

2.2.2 Gender difference in FIT performance

Most screening programs uses gender-uniform screening, i.e., men and women are screened in the same way. However, since 2015 the Stockholm-Gotland screening program applies a gender-specific strategy with lower cut-off levels in women, and a similar screening strategy is being launched from 2019 in Finland (126, 127). Kapidzic et al investigated gender differences in repeated screening rounds of a FIT-positive cohort (OC Sensor, cut-off level 10

µg/g in men and women). There was a higher positivity rate in men both at prevalent and incident rounds (11% in men and 6% in women, and 7% in men and 5% in women respectively), and these differences were evident also at higher cut-off levels. In men, a higher proportion of AN was detected than in women, but the gender-differences in CRC detection were non-significant. The PPV for AN – i.e., the proportion of AN among FIT positives compliant to colonoscopy, was equal in men and women. The authors concluded that FIT performed equal in men and women and the difference in detection rate was due to the higher prevalence of AN in men. In this study men had a higher false positive rate – i.e. those without AN at colonoscopy divided by the number of screened, possibly related to a higher number of non-AA that were FIT positive (128).

In the FIT arm of COLONPREV study a single FIT sample was used (OC sensor, cut-off 15 µg/g), which rendered a positivity rate of 7,2%. The positivity rate and the CRC and AA detection rate were higher in men than in women and in older individuals compared to younger. Raising the cut-off level from 15 to 40 µg/g decreased the detection of CRC presumably in older men but not in women. The detection of AA decreased in all groups at higher cut-off levels, but proportionally more among men (129).

The BLiTz study reported a higher sensitivity and lower specificity, and a higher PPV and lower NPV (those without AN among FIT-negatives) for AN in men than in women (130). A more recent study from the BLiTz cohort demonstrated similar results throughout a range of different FIT brands (131). A higher sensitivity, lower specificity and a higher PPV for AN in men than in women was also supported by findings in the COCOS trial, but participation rate was low and the COCOS study was not powered to detect gender differences (124).

In a recent study from the Scottish bowel screening program (cut-off 80µg/g) the authors investigated FIT levels in SD-CRCs by gender and found a significantly higher median FIT in men than in women in early stage and left-sided CRCs in addition to a higher positivity rate in men, and advocated a pilot study to evaluate lowering the cut-off level in women to 50 µg/g (132).

2.2.3 Risk factors for false positive and false negative test

A meta-analysis of 14 screening studies found that intake of nonsteroidal anti-inflammatory drugs (NSAID) -but surprisingly not anticoagulant therapy or acetyl salicylic acid (ASA), was associated with a false positive FIT, as NSAID increases the risk of gastrointestinal bleeding. A higher risk of a false negative FIT was seen in participants with family history of CRC, metabolic syndrome, advanced age, in men and smokers - factors that conveys a higher risk of developing CRC and in line with the previously mentioned relation between predictive values and prevalence (133). A more recent study of 4,600 participants in the BLiTz trial confirmed the increased risk for a false positive test in aspirin users, but also for participants with obesity, newly diagnosed IBD, old age, and in men and smokers, which could be due to other sources of bleeding associated with these conditions such as upper gastrointestinal cancers, peptic ulcers and inflammation (134). The study by Stegeman et al from the COCOS

trial not included in the meta-analysis reported that smoking and advanced age were risk factors for a false negative FIT, both of which are risk factors for CRC, and an increased risk of a false positive test was also seen in men, smokers and NSAID users (135). Furthermore, a randomized controlled study of a single dose aspirin prior to FIT screening did not prove an increased sensitivity as compared to the control group (136).

2.2.4 Risk stratification

Because of the large workload of colonoscopies generated in screening, several attempts have been made to create risk stratification models that increase the accuracy and select the high-risk participants for colonoscopy. FIT-based tailored screening could consider e.g., age, gender, the previous FIT-result, and family history of CRC, thereby creating subgroups of individuals with high- or low-risk for harboring AN and tailor them to different start- and stopping age of screening, different FIT cut-off levels or different re-screening intervals. Tailored screening could also aim at increasing the sensitivity of the test in different subgroups. However, it is important that the participation rate is not impaired if complex screening algorithms are implemented (137).

Auge et al analyzed 3,100 participants in the Barcelona screening program (OC sensor, cut-off 20 µg/g). PPV increased with age and was higher in men than in women, hence a risk stratification model was constructed including these variables that performed better than FIT measures alone in predicting AN (138). Stegeman et al evaluated data from the COCOS trial, and included age, calcium intake, CRC family history and smoking (but not gender) which provided a better prediction of AN than only FIT (139). A more recent study from the same research group included age and gender along with the FIT level to obtain the same risk of detecting AN at screening colonoscopy across groups, which rendered a limited improvement in sensitivity (140).

Omata et al investigated FIT performance (OC Micro, Eiken) in 1,100 asymptomatic participants of a health check-up at a tertiary hospital in Tokyo, of whom 70% were men and some at high-risk of developing CRC. Apart from FIT, higher BMI and age as well as male gender increased the risk of AN, and a nomogram was created with these variables that better predicted detection of AN than FIT alone (141).

Park et al performed a retrospective study of 3,700 participants in FIT screening in Korea that had undergone colonoscopy. Age, smoking, and FIT was independently associated with AN and was included in a risk stratification model along with diabetes and gender and displayed a better discriminatory ability for AN than only FIT. For CRC prediction, age and FIT was included in the model (142).

In a pilot study from England (OC sensor Eiken, cut-off 20µg/g) Cooper et al combined FIT with age, gender and screening history in a prediction model that increased the precision in selecting participants with AN (143).

2.3 Complications to screening colonoscopy

Colonoscopy is an invasive investigation that could cause complications such as bowel perforation, bleeding, and mortality. The procedure itself could be associated with discomfort or pain and also requires careful bowel preparation that has potential side-effects of hypotension and deranged electrolytes in certain patient groups (37, 38). Screening is being implemented worldwide and it is therefore important to determine the magnitude of adverse events in FIT screening cohorts.

In Denmark national population-based FIT screening was launched in 2014 for residents aged 50-74. Mikkelsen et al evaluated the first year of the screening program and demonstrated an overall complication rate of 0.61% per screening colonoscopy; 1.15% among participants having had a polypectomy and 0.14% in those who had not. None of the deaths were related to the colonoscopy. Moreover, the complications were underreported in the screening register when compared to the medical records (144).

Kooyker et al evaluated the mortality in the Dutch screening program 2014-2017 and the pilot study from 2013 and likewise concluded that complications in the endoscopy registries were underreported. The authors estimated 0.89 colonoscopy-related deaths per 10,000 screening colonoscopies from the causes of death register, and an excess death rate in those who underwent screening colonoscopy as compared to FIT negatives of 0.91 per 10,000, but medical records were not reviewed (145).

In the Veneto region in Italy, screening colonoscopy complications in 2004-2014 were estimated from hospital records and causes of death registers to 0.42% per screening colonoscopy; 0.64% in those having had a polypectomy and 0.14% in those with a diagnostic colonoscopy. The mortality rate was 1.24 per 10,000 colonoscopies (146).

2.4 Interval cancer in screening programs

A meta-analysis of the IC incidence was conducted on 17 FIT screening studies up to 2017 with varying cut-off levels. The pooled estimate for FIT IC was 20 (14-29) per 100,000 person-years of follow-up, and 15 (8-30) per 100,000 in the seven high-quality graded studies - as high heterogeneity (I^2=99%) was observed in the pooled data. Taking both FIT and gFOBT studies into account, IC were more common in the first as compared to third screening round and in older as compared to younger participants (147).

Two recent studies from different regions in Italy assessed the IC incidence of FIT screening programs by calculating the proportional IC incidence from the expected incidence had screening not been initiated, based on the background incidence from the period preceding screening. However, there is no comprehensive cancer register in Italy, so diagnoses also relied on hospital records. Zorzi et al demonstrated from five screening rounds in the Veneto region an IC incidence rate of 1.9 per 10,000 person-years. Sensitivity, calculated as 1 minus the proportional incidence, was 86.3%; higher in males than in females (89% vs 82%) and in distal as compared to proximal colon (94% vs 75%). The test sensitivity, calculated as SD-

CRC divided by total number of CRCs was 84% with similar differences between gender and colonic localization (148). Mancini et al in the Romagna area 2005-2012 calculated a proportional incidence of 0.06 and 0.21 in men and 0.17 and 0.28 in women for the first and second interval year respectively, and results remained after adjusting for selection bias of healthier subjects attending screening (149).

Toes-Zoutendijk et al evaluated the first screening round in the Dutch screening program with cut-off levels 15µg in the early and 47µg in the late study period, and the IC rate was 9.5 and 13.8 per 10,000 FIT-negatives respectively after adjusting for age. Test sensitivity was higher in men than in women (87% vs 83%) (150).

The Korean screening program with annual FIT was evaluated by Lee et al with regards to colonoscopy ICs occurring 6-60 months from a screening colonoscopy after a positive FIT in 2005-2010. The colonoscopy IC rate was 0.49 per 1,000 person-years and was higher in men than in women and increased with age. However, the colonoscopy compliance rate was 28%, and the time frame defining IC differed from other studies which affects the generalizability (151). In the Taiwanese national FIT screening program, a Colonoscopy IC after positive FIT was defined as CRC occurring within 0.5-3 years after the screening colonoscopy in those with AA, within 5 years in those with non-AA and within 10 years after a normal screening colonoscopy. An evaluation from 2004-2009 revealed a total IC rate of 1.14 per 1,000 person-years, and the risk of colonoscopy IC increased with age and FIT level (152).

2.5 Incidence and mortality in screening programs

The aim of screening is to reduce the CRC mortality, but the incidence of CRC will also be affected because of adenoma removal and polyp surveillance. In the National Polyp Study on surveillance of colonoscopy referral patients, there was a 76-90% decrease in CRC incidence after polypectomy and a 53% mortality reduction at the long-term follow-up (153, 154). In the Telemark Polyp Study a 60% decrease in CRC incidence and a non-significant decrease in mortality was demonstrated in the screening population after polypectomy (155).

In a recent evaluation of the Dutch screening program that commenced in 2014, the CRC incidence increased from 214 to 259 per 100,000 after implementation, and thereafter declined to 182 per 100,000 by 2019. Furthermore, there was an age-standardized decrease in CRC mortality from 88 per 100,000 prior screening to 65 per 100,000 by the year 2019 (156).

In California organized CRC screening with different modalities was implemented in 2000-2015. Participation rate improved during the period and the CRC incidence increased from 96 to 118 per 100,000 in the middle of the study period and then decreased to 71 per 100,000 in the late study period. The observed age-standardized CRC mortality decreased from 31 to 15 per 100,000 and the detection of advanced-stage CRC decreased with 36% (157). In the Danish screening program an even higher increase in incidence was seen; from 170 to 340 per 100,000 among invited as compared to the non-invited (158). An evaluation of the Florence FIT screening program 1993-99 revealed an increase in CRC incidence in

participants relative to non-participants the first 6-7 years after implementation, and thereafter a decline vs the non-attenders (159).

In the Taiwanese screening program 1.1 million participants were followed up to 6 years and compared to those non-exposed to screening. Despite the short follow-up time there was a 10% reduction in CRC mortality in the former population as compared to the latter (160). Moreover, among the non-compliers to screening colonoscopy the FIT level was positively associated to the adjusted CRC mortality (161).

In the Veneto region in Italy there was a peak incidence in CRC after screening initiation, and already 4 years later a 22% observed reduction in CRC mortality as compared to the preceding period, probably due to detection at an earlier stage. However, it is a remarkably fast decline, considering the moderate coverage, the paucity of SD-CRCs and a five-years survival rate of approximately 60% (162). In the Basque screening program there was an increased CRC incidence of 1-5% per year after screening implementation, and a decrease in mortality of 0.1-4% per year as compared to the standard population after a median follow-up of 4.6 years (163). In 11 Spanish regions with screening the CRC incidence rose with 10% after two years and the age-standardized CRC mortality decreased with 9% at 7 years as compared to 36 regions without screening (164).

2.6   Cost-efficiency in screening

The cost-efficiency of CRC screening has been assessed in several studies, RCTs, modelling studies and meta-analyses, and often compares the cost per Quality Adjusted Life-years (QALY) gained between different modes of CRC screening or compared to no screening. Cost-efficiency, e.g. $\leq$1-3 times the BNP per capita per QALY gained, has been demonstrated regardless of screening modality, the most cost-efficient being screening with colonoscopy or FIT (89).

A meta-analysis by Zong et al compared the cost-efficiency in biennial and annual FIT vs colonoscopy one-time and every 10th year in 23 studies and concluded that the FIT strategies were more cost-efficient, or cost-saving as compared to colonoscopy every 10th year in most of the studies (adopting a threshold of 50,000$ per QALY gained) (165). Included in the meta-analysis was a Swedish simulation study modelled on data from SCREESCO that conversely concluded that colonoscopy was more cost-efficient than FIT (166). Areia et al modelled screening with biennial FIT and colonoscopy every 10th year vs no screening in a Portuguese setting and estimated the cost per QALYs gained to 2,700€ and 48,300€ for FIT and colonoscopy screening respectively, the colonoscopy screening being above the threshold of 39,760€ /QALY defined as efficient (167).

Meulen et al simulated 480 gender-specific strategies vs gender-uniform FIT screening strategies varying different screening starting and stopping age, screening intervals and cut-off levels in men and women. The model was based on detection rates and positivity rates of a previous randomized trial (COREO-1). FIT screening was estimated to be less efficient

(QALY gained) and more costly for women and consequently the cost-efficiency was higher in men, but gender-specific screening was not more cost-efficient (168).

In the UK, FIT screening was implemented in 2019 with cut-off 120µg/g from the age of 60. A cost-efficiency study simulated reducing the screening start in men to 56 years (as the cumulative CRC incidence is similar in 56-year-old men as in 60-year-old women) and estimated that this would be more cost-efficient than screening everyone from the age of 58, with the same amount of screening resources used (169).

3 RESEARCH AIMS

The aim of this thesis was to increase the knowledge of FIT performance and to explore gender-specific screening with regards to CRC detection, interval cancers and screening costs in a Swedish FIT screening setting.

Research Questions:

1. How many FIT samples and what cut-off level should be used in a Swedish screening population?

2. What adenomas are detected and missed by FIT in a Swedish screening population?

3. Should different cut-off levels be used in men and women with regards to colonoscopy findings and screening costs?

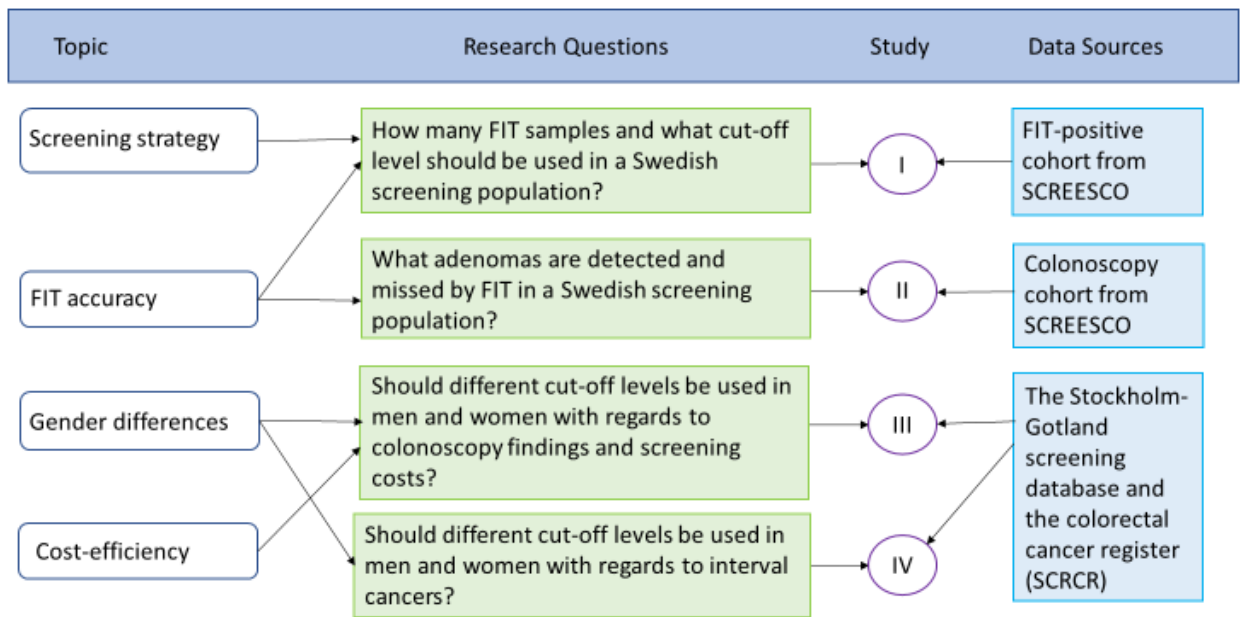4. Should different cut-off levels be used in men and women with regards to interval cancers?

Figure 4. Framework of the thesis.

4 MATERIALS AND METHODS

4.1 Study population

In both paper I and II the study population was derived from cohorts in the SCREESCO study. The SCREESCO study is an ongoing randomized controlled study on CRC mortality comparing screening with two rounds of FIT vs primary colonoscopy to no screening. No exclusions were applied except a previous CRC diagnosis or having had a proctocolectomy. In the colonoscopy and FIT arms, 30,500 and 60,000 60-years-olds were invited respectively, and 183,000 randomized non-invited individuals served as controls. The FIT arm applied two samples with a cut-off level of 10µg/g (OC Sensor Eiken, Japan), and when at least one of the two samples were positive the participant was offered colonoscopy. The study covered all regions in Sweden except the Stockholm-Gotland-region that has an ongoing screening program and the county of Västernorrland, who declined inclusion in the study. The inclusion period was 2014-2020. Results on the primary outcome is estimated at 2034 (170).

In **paper I** the study population consisted of individuals invited to the FIT arm of SCREESCO between March 2014 to Aug 2015 who underwent a complete bowel investigation before Dec 2015.

In **paper II** the study population consisted of participants in the FICO (FIT Colonoscopy) trial of SCREESCO. In FICO, 1155 participants of the colonoscopy arm were also invited to provide two FIT samples prior to the bowel investigation to assess colonoscopy findings in the FIT-negatives. The inclusion period was March 2016 to Feb 2017.

In paper III and IV the study population was derived from the Stockholm-Gotland screening program that since 2008 invites 60–69-year-olds to biennial FOBT screening. All residents are covered in the invitations and no exclusions are applied except those who were referred to polyp surveillance from previous screening rounds. In 2015 the screening program shifted to FIT with gender-specific cut-off levels for a positive test; 40µg/g in women and 80µg/g in men. The positives were referred to colonoscopy at the closest participating endoscopy unit.

In **paper III** the study population consisted of all invited to screening Oct 2015 to Dec 2017 that had completed FIT analysis within one month from sampling and the bowel investigation within 6 months. If several invitations occurred during the study period the first or the first complete participation was included, thus every participant was only counted once.

In **paper IV** the study population included those invited to the first screening round with FIT, hence from Oct 2015 to Sept 2017. The follow-up period was two years with regards to CRC diagnosis.

4.2 Data sources

In **paper I** the SCREESCO screening register was used which included information on the FIT results and date of sampling, colonoscopy findings, colonoscopy quality parameters, measured Body Mass Index (BMI) and a questionnaire completed by the participant at the

endoscopy unit on bowel habits and medication. To allow for a detailed analysis of adenoma and CRC characteristics all adenoma and CRC findings were verified against the pathology reports provided by the local endoscopy centers.

In **paper II** the FIT results and dates were provided directly from the study laboratory Aleris Medilab, otherwise as described above.

In **paper III and IV** information on FIT results and dates, colonoscopy findings and colonoscopy quality parameters were collected from the screening register at the Regional Cancer Center (RCC) in Stockholm. All CRCs were identified in the Swedish Colorectal Cancer Register (SCRCR) which has a completeness of 99% and a validity of >90%. The SCRCR commenced in 1995 for rectal cancer and in 2007 for colonic cancer and covers all adenocarcinomas in the large bowel (171).

Regarding the cost analysis in paper III, costs for invitation and FIT analysis were derived from the laboratory and invitation costs managed by RCC. The Nord-DRG register was used to estimate the cost of colonoscopy. The Nord-DRG is put together by the National Board of Health and Welfare and the Swedish regions and consists of weights for all groups of diagnoses and procedures within hospitals and clinics depending on the average cost per patient. The cost of the reference weight=1 is updated each year (172).

In paper IV the Cancer Register was used to calculate the experienced incidence rate (EIR) of CRC, i.e., the CRC incidence before screening implementation in the Stockholm-Gotland region. The Swedish Cancer register started in 1958 and is a national register that covers all diagnosed primary cancers, and reporting is mandatory for clinicians according to health legislation (173)

4.3 Statistical methods

In **Paper I** cut-off levels for the first sample, the highest of the two, the lowest of the two and for the mean of the two samples at cut-off levels 10, 15, 20, 40 and 80µg/g was assessed in relation to findings at colonoscopy. PPV for CRC and AA and the Number needed to Scope (NNS) per AA and CRC was calculated. PPV was defined as the number of CRC or AA among FIT positives and NNS as the inverse of the detection rate of CRC or AA among those who underwent colonoscopy.

The association between FIT level in the first sample and the different variables (CRC, AA, non-AA, other findings) were assessed with univariate and multivariable median regression analysis in all with a first FIT sample. The medication, gender, other findings and BMI variables were included for adjustment in the multivariable model. The association between the FIT level in the first FIT sample and adenoma characteristics (size, localization, grade of dysplasia, morphology, and number of adenomas) in those with adenomas and a first FIT sample was assessed in a similar way. Variables in the univariate analysis were presented with median and interquartile ranges of FIT and a 95% confidence interval (CI) for the

median difference. In the multivariable analysis a combined p-value was used for categorical variables with multiple categories. A p-value of <0.05 was considered statistically significant.

To compare the detection rate of CRC between two screening strategies with different cut off levels and number of samples, two different data sets were constructed, one for each of the strategies. Logistic regression analysis was used to compare the two strategies, and the standard error was corrected within the generalized estimating equations framework, since the same individual could be present in both data sets (174). The proportion of AN, localization of adenomas and CRC, and PPV in men vs women was assessed with Chi-squared test. The analyses were carried out in STATA v.13.

In **paper II** the first FIT sample, any of the two samples and the mean of two samples were evaluated at cut off levels 10, 20, 40 ,60 and 80 µg/g. As the study cohort consisted of participants invited to both colonoscopy and FIT sampling the sensitivity, specificity and the NPV for advanced neoplasia was possible to calculate. The PPV was defined as the proportion of AN among FIT positives and the NPV as the proportion of participants without AN among FIT negatives. The 95% CI for these measures and for the positivity rate was calculated with the Clopper Pearson method. The difference in sensitivity and specificity between different screening strategies were assessed with McNemar test. The false negative rate was calculated as the proportion of negatives in individuals with AN (1-sensitivity). The false positive rate was calculated as the proportion of positives in individuals without AN (1-specificity).

In those with adenoma, a univariate analysis was carried out on the possible association between FIT positivity (any of the two samples ≥10µg/g) and adenoma characteristics (localization, villosity, shape, size, grade of dysplasia and gender) with Chi squared or McNemar test. The total number of adenomas and FIT positivity was assessed with Mann-Whitney's U test. High-risk dysplasia was defined as HGD or dysplasia in an SSA. Multivariable logistic regression was conducted to model the ORs with 95% CI for the association of FIT positivity and the above variables of adenoma characteristics. The categorical variables having more than two categories were assessed together with a Wald test. A sensitivity analysis was then carried out excluding those on ASA and NSAID medication. All analyses were carried out in STATA v.13.

In **paper III** the FIT positivity was defined as the number of individuals with FIT above or equal to the cut-off level divided by the number of individuals with analyzable results. AA was defined as high-risk adenomas that required a follow-up colonoscopy. PPV was calculated as the number of participants with CRC or AA among the FIT-positives who underwent colonoscopy and was estimated overall and separately for men and women and at different cut-off levels and presented with 95% CI. NNS was defined as the number needed to undergo screening colonoscopy per detected CRC or AA and estimated overall and in subgroups as above. In women, FIT was further categorized in 40-79µg/g and ≥80µg/g and differences in FIT category and CRC and AA proportion, CRC stage and localization was assessed with Chi-squared test. Differences in men vs women in the proportion of CRC, AA

and normal colonoscopy, CRC stage and localization was also assessed with Chi-squared test. A p-value of <0.05 was considered statistically significant. The statistical analyses were done in R version 3.6.2.

For the estimates of the screening costs, the current strategy was compared to having a cut-off level of 80μg/g in both men and women. A separate sensitivity analysis was carried out excluding the cost for follow-up colonoscopies.

In **paper IV** the participation rate was defined as the proportion of individuals with a valid FIT among the invited individuals, calculated overall and in subgroups. The FIT positivity rate was defined as the number of individuals with FIT above or equal to the cut-off level divided by the number of individuals with analyzable results, and colonoscopy compliance defined as having had a screening colonoscopy after a positive FIT. The rates differences in subgroups were assessed with Chi-squared test. The PPV was defined as the number of individuals with SD-CRC divided by the number of FIT positives and presented with 95% CI. Differences in PPV between subgroups were assessed as above.

The IC rate was defined as the number of total ICs per 10,000 negatively screened -either negative FIT or positive FIT and negative screening colonoscopy, and was calculated overall for the total screening round of two years and separately by year after invitation. The number of negatives for each year after invitation was assumed to be half of the total number of negatives for the screening round. The IC incidence rate was calculated as the number of ICs per 100,000 person-years of follow-up. The follow-up period was two years for every individual with regards to CRC diagnosis, except for those who were diagnosed with an IC within the first year after invitation. The test sensitivity was calculated as the proportion of SD-CRC among all CRCs. The experienced incidence rate (EIR) was calculated as the mean incidence per 100,000 in each age and gender groups for the ten years preceding screening implementation (1998-2007) in the Stockholm-Gotland region. The ratio of the IC incidence rate and the EIR was calculated stratified by each age and gender group for the total screening round of two years and by each year from invitation. The 95% CI of the rate ratio was calculated with the exact Poisson method.

When estimating the IC rate, IC incidence rate and test sensitivity had the cut-off levels been 80μg/g in both men and women, it was assumed that in women with FIT 40-79μg/g all SD-CRCs, colonoscopy-ICs and CRCs in those non-compliant to colonoscopy would classify as FIT-ICs.

The Chi-squared test was used to analyze the differences in test sensitivity and IC rate between subgroups, and differences in proportion of CRC characteristics between SD-CRC and FIT ICs and CRCs in non-participants, and a p-value <0.05 was considered statistically significant.

4.4 Ethical considerations

There are several ethical issues to take into account with regard to colorectal cancer screening and screening in general.

Firstly, screening involves healthy individuals, that are invited to screening, and who have not spontaneously contacted care facilities because of worries or (bowel) symptoms. Colonoscopy confers, as discussed above, a small risk of serious complications. It is likely that a patient with symptoms is more inclined to taking medical risks for the sake of the investigation and treatment of a disease as compared to a healthy individual participating in a disease prevention program. Therefore, it is essential to analyze the consequences of screening in the population and to minimize the adverse events so that the gain in health outweighs the risks. We consider the risk of colonoscopy complications to be very small in relation to the potential of screening to reduce disease mortality. Nevertheless, it is of utter importance that the screening invitees are aware of the risks to be able to make an informed decision on participation.

Secondly, a positive FIT could lead to emotional distress among participants before undergoing the colonoscopy to identify the source of bleeding. Some tests are false positive, and the colonoscopy is normal. In most of the participants no cancer is detected. Is it justifiable to alarm people with a positive test when the majority of colonoscopies do not detect any serious illness? This is an ethical dilemma in screening for rare diseases when the diagnosis in most cases could be dismissed. Again, it is important that invitees are informed about what happens after a positive test and that it can be positive for other reason than cancer. The central screening organization includes screening nurses that can be contacted for further information.

A third aspect of screening is ensuring that participants with a positive FIT comply with the full bowel investigation. FIT positives comprise a high-risk group and are identified as such by the screening, thereby the screening organization is partly responsible for an accurate follow-up (161). Even though participation and compliance are voluntarily, it is important with thorough information and reminders to complete the investigation. As in general health care, compliance to treatments and follow-up is sometimes a challenge for which both the patient and the health care provider is responsible.

A fourth aspect of screening is to prioritize the limited resources for health care, e.g., the demand for costly colonoscopy facilities required for screening. Furthermore, Swedish health care is obliged to be equal and to give priority to the most severely diseased. Participants of screening are often from socio-economic strong groups. How should health care providers prioritize between health care for sick patients and preventive tasks for healthy subjects? Swedish health care is government-funded and has large economical resources. Moreover, in the long run screening could be cost saving because disease mortality is decreased, and an early diagnosis leads to less extensive and less costly treatment. However, screening is only

tenable if the uptake is high and equally distributed, and efforts should be made to evaluate and increase participation in socio-economic weak groups.

The ethical considerations specific to this project is mainly that of confidentiality and personal integrity. We handle delicate personal information from the study database, the screening register and SCRCR of which public distribution or indiscrete management would be a violation of the personal integrity. The personal data are handled according to the GDPR legislation and participation is voluntarily. As much as possible, we use non-identifiable data to minimize the risk of violation of personal integrity. The information is managed on a group level and the number of study subjects is very large, hence specific individuals are not traceable in the data presentation. The research group uses ELN which enables transparency, trackability and a safe management of delicate information. The studies have received ethical permission by the reginal ethics board, registration number 2012/2038-31 and 2019-04850.

5 RESULTS

5.1 Paper I

In the study cohort, 12 383 had an analyzable FIT, 1,396 were FIT positive and 1,182 underwent a complete investigation. Sixty-one participants were excluded from analyses regarding the first FIT sample since they had the same date on both samples, and four had only one valid FIT and were thus excluded from analyses of the mean of two samples. Median of the first, the highest and the mean of two FIT samples were 15.8 (4.4-41.8), 30 (15.2-75.8) and 18.4 (10-46) µg/g respectively. Caecal intubation, clean bowel and >6 minutes withdrawal time was accomplished in 95-96%. Details of the 1,182 participants are listed in Table 2.

Table 2. Basic variables in 1,182 FIT screening participants.

Colonoscopy findings	Participants, n
All	1182
Women	551
Men	631
BMI, mean	27.5
Acetylsalicylic acid, yes	216
weekly dose, n (range)	(1-20)
NSAID yes	183
weekly dose, n (range)	(1-20)
CRC	27†
Proximal	4
Distal	23
Adenoma	490†
Proximal only	116
Distal only	279
Both	95
SSA/P	37
No dysplasia	21
LGD	421
Tubular	299
<5mm	136
5-9mm	165
Number of adenomas/participant	
(range)	(0-25)
1	288
2	117
3	45
≥4	40
AA	269†
10-19mm	148
≥20mm	41
HGD	48
Villous or tubulovillous	148
Proximal only	47
Distal only	152
Non-AA	221†
Other findings	439†
Diverticular disease	493
Inflammation	56
Hemmoroids	237
Angiodysplasia	24
Normal colonoscopy	226†

Proximal= caecum to splenic flexure. AA= Advanced adenoma (≥10mm, adenomas with high grade dysplasia or a villous component, or ≥3 tubular adenomas <10mm). Non-AA= Non advanced adenoma (≤2 tubular adenomas <10mm). SSA/P= sessile serrated adenoma, or hyperplastic polyp >9mm. HGD= high grade dysplasia. LGD= low grade dysplasia. Other findings include diverticulas, inflammation, hemmoroids or angiodysplasia.   †) According to most advanced lesion.

CRC was detected in 27 (2.3%) and AA in 269 (23%) of the participants and the median of their first FIT sample was significantly higher as compared to in those with non-AA, other findings, or normal colonoscopy at multivariable analysis (Table 3).

Table 3. Multivariable analysis of first FIT sample level in 1,115 screening participants.

Variable	Coeff.	p-value	95% CI	Combined p-value
NSAID	0.19	0.332	(-0.20; 0.58)	
Acetylsalicylic acid	-0.07	0.825	(-0.69; 0.55)	
BMI	0.10	0.487	(-0.19; 0.40)	
Colonoscopy findings				
Normal	Ref.			0.005
Other	1.89	0.442	(-2.93; 6.70)	
Non-AA	2.62	0.254	(-1.88; 7.11)	
AA	10.5	0.002	(4.02; 17.0)	
CRC	224	0.035	(15.6; 432)	
Gender				
Female	Ref.			
Male	-1.50	0.261	(-4.12; 1.12)	
Hemorroids				
No	Ref.			
Yes	-0.88	0.622	(-4.39; 2.63)	
Inflammation				
No	Ref.			
Yes	-0.28	0.941	(-7.66; 7.10)	
Diverticulas				
No	Ref.			
Yes	-1.86	0.319	(-5.52; 1.80)	
Angiodysplasia				
No	Ref.			
Yes	4.19	0.824	(-32.8; 41.2)	

NSAID= Non-steroidal anti-inflammatory drugs. Acetylsalicylic acid, weekly dose. AA= Advanced adenoma (≥10mm, adenomas with high grade dysplasia or a villous component, or ≥3 tubular adenomas <10mm). Non-AA= Non advanced adenoma (≤2 tubular adenomas <10mm).

In the 449 participants with any adenoma the median first FIT sample was independently associated to adenoma size regardless of the number of adenomas, medication, histology, localization, BMI, gender and other findings (Table 4).

Table 4. Multivariable analysis of adenoma characteristics and first FIT sample level in 449 screening participants with adenoma.

Variable	Coeff.	p-value	95% CI	Combined p-value
Number of adenomas†	1.07	0.648	(-3.53; 5.66)	
NSAID	-0.81	0.055	(-1.63; 0.02)	
ASA	-1.12	0.004	(-1.86; -0.37)	
BMI	0.48	0.084	(-0.06; 1.03)	
Growth pattern				
Tubular	Ref.			0.679
Tubulovillous or villous	-4.05	0.463	(-14.9; 6.80)	
Other (SSA/HP)	6.19	0.651	(-20.7; 33.1)	
Dysplasia				
LGD	Ref.			0.344
HGD	-0.55	0.961	(-22.7; 21.6)	
No dysplasia	-21.1	0.144	(-49.6; 7.26)	
Localization				
Proximal	Ref.			0.657
Distal	1.28	0.681	(-4.84; 7.41)	
Both	3.93	0.373	(-4.73; 12.6)	
Size				
<5mm	Ref.			0.038
5-9mm	3.57	0.324	(-3.54; 10.7)	
10-19mm	14.7	0.007	(3.97; 25.4)	
≥20mm	26.7	0.113	(-6.32; 59.7)	
Gender				
Female	Ref.			
Male	0.46	0.861	(-4.75; 5.68)	
Hemorroids				
No	Ref.			
Yes	1.70	0.533	(-3.65; 7.04)	
Inflammation				
No	Ref.			
Yes	3.19	0.757	(-17.1; 23.4)	
Diverticulas				
No	Ref.			
Yes	-2.45	0.355	(-7.67; 2.76)	
Angiodysplasia				
No	Ref.			
Yes	34.8	0.213	(-20.0; 89.6)	

ASA= Acetylsalicylic acid, weekly dose. SSA/HP= Sessile Serrated Adenoma/Hyperplastic polyp >9mm. LGD= low grade dysplasia. HGD= high grade dysplasia. Proximal= caecum to splenic flexure. † Regardless of adenoma size.

In Table 5 the FIT positivity rate, colonoscopy findings, NNS and PPV are given for the first, the highest of two, the mean of two and the lowest of two samples at different cut-off levels. At each of the cut-off levels, the CRC and AA detection increased when using the mean of two samples or the highest of two samples as compared to the first sample. If the cut-off level was raised from 10 to 40µg/g, 19-26% of the CRCs and 44-49% of the ANs would have been missed due to negative test depending on the number of samples used. Correspondingly, the number of colonoscopies required would decrease with 59-66%. Taking both the cut-off level and the number of samples into account, the CRC detection rate was significantly higher with first sample at cut-off 20 or 40µg/g as compared to mean of two samples at cut-off level 40 or 80µg/g (p-value 0.006 and 0.003 respectively). The PPV for CRC and AA increased with the cut-off level and was higher in the first sample as compared to mean of two or any of the two samples.

Table 5. FIT positivity, colonoscopy findings and positive predictive value for 12,383 screening participants at different cut-off levels and number of samples.

Test (cut off µg Hb /g)sample	FIT positive (%)	Colonoscopies	CRC (%)	NNS, CRC	PPV CRC %	AA (%)	NNS, AA	PPV AA %	Only prox AA (%)	Non-AA (%)	Other (%)	Normal (%)
FIT(10)first	887 (7.5)	750	25 (3.3)	30	2.8	180 (24)	4.2	20	33 (4.4)	137 (18)	270 (36)	138 (18)
FIT(15)first	691 (5.8)	579	24 (4.1)	24	3.5	154 (27)	3.8	22	25 (4.3)	107 (18)	196 (34)	98 (17)
FIT(20)first	571 (4.8)	481	22 (4.6)	22	3.9	140 (29)	3.4	25	22 (4.6)	82 (17)	160 (33)	77 (16)
FIT(40)first	347 (2.9)	292	19 (6.7)	15	5.5	94 (32)	3.1	27	11 (3.8)	39 (13)	98 (34)	42 (14)
FIT(80)first	208 (1.8)	176	17 (9.7)	10	8.2	56 (32)	3.1	27	5 (2.8)	18 (10)	61 (35)	24 (14)
FIT (10) ≥1 of 2 pos	1396 (11)	1182	27 (2.3)	44	1.9	269 (23)	4.4	19	47 (4.0)	221 (19)	439 (37)	226 (19)
FIT (15) ≥1 of 2 pos	1071 (8.6)	901	25 (2.8)	36	2.3	221 (25)	4.1	21	35 (3.9)	170 (19)	321 (36)	164 (18)
FIT (20) ≥1 of 2 pos	890 (7.2)	748	24 (3.2)	31	2.7	199 (27)	3.8	22	29 (3.9)	133 (18)	258 (34)	134 (18)
FIT (40) ≥1 of 2 pos	572 (4.6)	483	20 (4.1)	24	3.5	147 (30)	3.3	26	18 (3.7)	77 (16)	170 (35)	69 (14)
FIT (80) ≥1 of 2 pos	332 (2.7)	282	18 (6.4)	16	5.4	94 (33)	3	28	8 (2.8)	35 (12)	99 (35)	36 (13)
FIT(10)mean of 2	1046 (8.5)	887	26 (2.9)	34	2.5	220 (25)	4.0	21	37 (4.2)	167 (19)	308 (35)	166 (19)
FIT(15)mean of 2	794 (6.4)	669	23 (3.4)	29	2.9	181 (27)	3.7	23	28 (4.2)	114 (17)	238 (36)	113 (17)
FIT(20)mean of 2	649 (5.2)	550	21 (3.8)	26	3.2	161 (29)	3.4	25	24 (4.4)	90 (16)	196 (36)	82 (15)
FIT(40)mean of 2	390 (3.2)	331	20 (6.0)	17	5.1	106 (32)	3.1	27	14 (4.2)	47 (14)	112 (34)	46 (14)
FIT(80)mean of 2	228 (1.8)	198	17 (8.6)	12	7.5	70 (35)	2.8	31	4 (2.0)	22 (11)	69 (35)	20 (10)
FIT(10)2 of 2 pos	447 (3.6)	374	21 (5.6)	18	4.7	116 (31)	3.2	26	21 (5.6)	62 (17)	124 (33)	51 (14)
FIT(15)2 of 2 pos	342 (2.8)	279	20 (7.2)	14	5.8	95 (34)	2.9	28	15 (5.4)	41 (15)	93 (33)	30 (11)
FIT(20)2 of 2 pos	285 (2.3)	228	19 (8.3)	12	6.7	86 (38)	2.7	30	14 (6.1)	35 (15)	65 (29)	23 (10)
FIT(40)2 of 2 pos	157 (1.3)	128	17 (13)	7.5	11	56 (44)	2.3	36	6 (4.7)	11 (8.6)	34 (27)	10 (7.8)
FIT(80)2 of 2 pos	91 (0.7)	76	14 (18)	5.4	15	34 (45)	2.2	37	2 (2.6)	3 (3.9)	19 (25)	6 (7.9)

First = the first of two samples. ≥1 of 2 pos = at least one of two samples above cut-off. Mean of 2 = mean of two samples above cut-off. 2 of 2 pos= both samples above cut-off. Of the 12 383 participants, 554 are excluded from first sample as they have the same date on both tests, and 9 are excluded from mean of two samples and both samples above cut-off as they have only one valid test. Of the colonoscopies, 61 are excluded from first sample as they have the same date on both tests, and 4 are excluded from mean of two samples and both samples above cut-off as they have only one valid test. The number of colonoscopies required for each cut off is based on the FIT results of the 1182 colonoscopy participants.

The proportion of AN was significantly higher in men than in women (n=180 vs 116), but the proportion of proximal adenomas and CRC was equal between genders (n=64 vs 56, p-value 0.23). the PPV for AA was significantly higher in men than in women at all cut-off levels <40µg/g and for mean of two samples at cut-off 80µg/g, but equal between genders for CRC. Lowering the cut-off level in women from 40 to 20µg/g or 80 to 40µg/g for the first or the mean of two samples would generate the same AN detection as in men, but increase the colonoscopies required by 26-34%.

5.2 Paper II

In the FICO cohort, 806 completed the questionnaire and the investigation. In 48 of the participants a random sample of the two were used as the first FIT sample since they had the same date on both samples, and six had only one valid FIT and were thus excluded from analyses of the mean of two samples. Of the participants, 102 (12.7%) were FIT positive (any of the two samples ≥10µg/g). CRC was detected in 1 (0.1%), AA in 80 (9.9%) and non-AA in 145 (18%) (Table 6).

Table 6. Basic variables in 806 colonoscopy screening participants.

Variable	All participants (n=806), N (%)	Participants with FIT <10µg/g (n=704), N (%)	Participants with at least one of two FIT ≥10µg/g (n=102), N (%)
Gender			
Men	390 (48)	334 (47)	56 (55)
Women	416 (52)	370 (53)	46 (45)
NSAID			
Yes	49 (6.0)	43 (6.1)	6 (5.8)
No	750 (93)	656 (93)	94 (92)
Missing	7 (1.0)	5 (0.7)	2 (2.0)
ASA			
Yes	67 (8.3)	54 (7.7)	13 (13)§
No	735 (91)	647 (92)	88 (86)
Missing	4 (0.5)	3 (0.4)	1 (1.0)
BMI			
Median [range]	26.1 [17.0-44.3]	25.9 [17.9-42.7]	26.8 [17.0-44.3]
Missing	1 (0.1)	1 (0.1)	0 (0.0)
Caecal intubation.	781 (97)	684 (97)	97 (95)
Caecal withdrawal time >6min	781 (97)	682 (97)	99 (97)
Bowel preparation. Boston scale ≥2	790 (98)	693 (98)	97 (95)
Colonoscopy findings†			
CRC	1 (0.1)	0 (0.0)	1 (1.0)
AA	80 (9.9)	60 (8.5)	20 (20)
Non-AA	145 (18)	131 (19)	14 (14)
Other	230 (29)	197 (28)	33 (32)
Normal	350 (43)	316 (45)	34 (33)
Number of adenomas/participant			
[range]	[0-35]	[0-8]	[0-35]
1	160 (20)	136 (19)	24 (24)
2	42 (5.2)	36 (5.1)	6 (5.8)
≥3	23 (2.9)	19 (2.7)	4 (3.9)

AA (n=80)	Participants with AA, N=80 (%)	AA Participants with FIT <10µg/g, N=60 (%)	AA Participants with ≥1 of 2 FIT ≥10µg/g, N=20 (%)
Size <10mm	31 (39)	27 (87)	4 (13)
Localization			
Proximal	36 (45)	33 (92)	3 (8.3)
Distal	26 (33)	13 (50)	13 (50)
Both	18 (23)	14 (78)	4 (22)
Histology			
Tubular	19 (24)	14 (74)	5 (26)
SSA/HP no dysplasia	14 (18)	14 (100)	0 (0)
LGD	47 (59)	36 (77)	11 (23)
Shape			
Pedunculated	13 (16)	4 (31)	9 (69)
Flat or broad based	67 (84)	56 (84)	11 (16)
Gender			
Men	46 (58)	30 (65)	16 (35)
Women	34 (43)	30 (88)	4 (12)

NSAID=Non-steroidal anti-inflammatory drugs. ASA=Acetylsalisylic acid. Other= diverticular disease, hemmoroids, angiodysplasia or inflammation. FIT= Fecal immunochemical test. SSA/HP no dysplasia= sessile serrated adenoma/polyp or hyperplastic polyp ≥10mm without dysplasia. †) According to most advanced lesion.§) The difference in proportion of ASA medication between FIT positives and negatives is non-significant.

In Table 7 the FIT positivity rate, PPV, NPV, sensitivity and specificity for AN at different cut-off levels and number of samples is listed. Sensitivity and specificity ranged from 7-26% and 89-99% respectively, corresponding to a PPV of 21-52% and a NPV of 91-92%. The false negative rate was 74-93%. At each cut-off level, there was no gain in sensitivity with two samples as compared to one. Specificity was significantly higher with one sample than for any of the two at cut-off levels 10 and 20µg/g.

Table 7. Sensitivity and specificity for advanced neoplasia at different FIT cut-off levels and number of samples in 806 screening participants.

Test, cut off µg/g	Positivity (95% CI)	True positive	False positive	True negative	False negative	Sensitivity* (95% CI)	Specificity** (95% CI)	PPV (95% CI)	NPV (95% CI)
≥1 of 2†									
10	12.7 (10.4-15.0)	21	81	644	60	25.9 (16.8-36.9)	88.8 (86.3-91.0)	20.6 (13.2-29.7)	91.5 (89.2-93.4)
20	7.2 (5.4-9.0)	19	39	686	62	23.5 (14.8-34.2)	94.6 (92.7-96.1)	32.8 (21.0-46.3)	91.7 (89.5-93.6)
40	3.7 (2.4-5.0)	13	17	708	68	16.0 (8.80-25.9)	97.7 (96.3-98.6)	43.3 (25.5-62.6)	91.2 (89.0-93.1)
60	2.5 (1.4-3.6)	10	10	715	71	12.3 (6.10-21.5)	98.6 (97.5-99.3)	50.0 (27.2-72.8)	91.0 (88.7-92.9)
80	2.2 (1.2-3.3)	9	9	716	72	11.1 (5.20-20.0)	98.8 (97.7-99.4)	50.0 (27.2-72.8)	90.9 (88.6-92.8)
Mean‡									
10	9.6 (7.6-11.7)	20	57	662	61	24.7 (15.8-35.5)	92.1 (89.9-93.9)	26.0 (16.6-37.2)	91.6 (89.3-93.5)
20	4.6 (3.2-6.1)	14	23	696	67	17.3 (9.80-27.3)	96.8 (95.2-98.0)	37.8 (22.5-55.2)	91.2 (89.0-93.1)
40	2.6 (1.5-3.7)	11	10	709	70	13.6 (7.00-23.0)	98.6 (97.5-99.3)	52.4 (29.8-74.3)	91.0 (88.8-92.9)
60	2.1 (1.1-3.1)	8	9	710	73	9.90 (4.40-18.5)	98.7 (97.6-99.4)	47.1 (23.0-72.2)	90.7 (88.4-92.6)
80	1.8 (0.8-2.7)	6	8	711	75	7.40 (2.80-15.4)	98.9 (97.8-99.5)	42.9 (17.7-71.1)	90.5 (88.2-92.4)
First§									
10	8.6 (6.6-10.5)	16	53	672	65	19.8 (11.7-30.1)	92.7 (90.5-94.5)	23.2 (13.9-34.9)	91.2 (88.9-93.1)
20	4.2 (2.8-5.6)	12	22	703	69	14.8 (7.90-24.4)	97.0 (95.4-98.1)	35.3 (19.8-53.5)	91.1 (88.8-93.0)
40	2.4 (1.3-3.4)	8	11	714	73	9.90 (4.40-18.5)	98.5 (97.3-99.2)	42.1 (20.3-66.5)	90.7 (88.5-92.7)
60	1.6 (0.7-2.5)	6	7	718	75	7.40 (2.80-15.4)	99.0 (98.0-99.6)	46.2 (19.2-74.9)	90.5 (88.3-92.5)
80	1.6 (0.7-2.5)	6	7	718	75	7.40 (2.80-15.4)	99.0 (98.0-99.6)	46.2 (19.2-74.9)	90.5 (88.3-92.5)

Advanced neoplasia= adenocarcinoma or advanced adenoma (=adenoma ≥10mm or ≥3 <10mm or adenomas with villous growth or high grade dysplasia). †)At least one of two samples above or equal to cut-off. ‡)Mean of two samples above cut-off. 6 participants were excluded as they sent only one sample. §)First of two samples. In 48 participants both tests are from the same date, therefore a random sample was chosen as the first FIT. *)The differences in sensitivity for each cut-off level are non-significant. **)The differences in specificity for each cut-off level are non-significant except for ≥1 of 2 samples and First sample at cut-off 10 µg/g (p=0.02) and ≥1 of 2 samples and First sample at cut-off 20 µg/g (p=0.04).

In the 225 participants with any adenoma, pedunculated shape, high-risk dysplasia and male gender were independently associated with FIT positivity (Table 8). However, when restricting the analysis to those 198 without ASA or NSAID medication, only the association to FIT positivity between high-risk dysplasia and pedunculated shape remained.

Table 8. Odds ratio for FIT positivity (at least one of two samples ≥10µg Hemoglobin/g) in 225 screening participants with adenoma.

Variable	Odds ratio	95% CI	p-value
Number of adenomas†	1.12	0.94-1.32	0.209
Growth pattern			
HP/SSA/Tubular+SSA	1 (ref)		
Tubular	2.38	0.50-11.5	0.279
Tubulovillous or villous	7.35	1.36-39.7	0.020
Dysplasia			
Low-risk dysplasia	1 (ref)		
High-risk dysplasia	7.34	1.64-32.8	0.009
Localization			
Proximal	1 (ref)		
Distal	1.66	0.57-4.83	0.355
Both	0.86	0.16-4.47	0.853
Size			
<5mm	1 (ref)		
5-9mm	2.85	0.82-9.90	0.099
10-19mm	3.28	0.64-16.9	0.156
≥20mm	3.06	0.41-22.8	0.276
Gender			
Men	1 (ref)		
Women	0.35	0.13-0.93	0.036
Shape			
Flat/Broad based	1 (ref)		
Pedunculated	5.09	1.57-16.5	0.007

SSA/HP= Sessile Serrated Adenoma/Hyperplastic polyp >9mm. Tubular+SSA= Synchronous SSA and Tubular adenoma. Low-risk dysplasia= low grade dysplasia in an adenoma, or no dysplasia in SSA. High-risk dysplasia= high grade dysplasia in an adenoma, or dysplasia in SSA. † Regardless of adenoma size.

SSA was more common in women and compared to villous and tubulovillous growth less prone to be FIT positive. In men vs women sensitivity and specificity for AA was 34.8% vs 11.8% (p-value 0.021) and 88.4% vs 89.0% respectively (p-value 0.8) at cut-off ≥10µg/g for any of the two samples.

5.3 Paper III

During the study period, 229,944 were invited. Both participation and colonoscopy compliance were significantly higher in women than in men; 72% vs 65% and 90% vs 86% respectively, but positivity rate was equal (2.7%). Three individuals had symptomatic CRC diagnosed at the same time as participating in screening and were excluded. In the 3758 included colonoscopies, CRC was found in 138 (8.3%) men and 120 (5.8%) women (p-value 0.03) (Table 9). Proximal CRC was more common in women than in men (31% vs 18%, p-value 0.025), as were a normal investigation (24% vs 17%, p-value <0.05).

Table 9. Colonoscopy findings and quality parameters in 1672 men and 2086 women that completed screening.

Variables	Total, n (%)	Men, n (%)	Women, n (%)	p-value
Colonoscopy participants, n	3758 (100)	1672 (44)	2086 (56)	
Age, median (IQ range)	64 (62-67)	64 (62-67)	64 (62-68)	
Clean colon, yes	3634 (97)	1611 (96)	2023 (97)	
Caecum investigated, yes	3651 (97)	1633 (98)	2018 (97)	
≥10min withdrawal time	3733 (99)	1661 (99)	2072 (99)	
Bowel perforation, yes	5 (0.1)	2 (0.1)	3 (0.1)	
Other complications, yes	26 (0.7)	15 (0.9)	11 (0.5)	
Median fecal Hemoglobin, µg/g (IQ range)	140 (81-359)	218 (121-534.2)	94 (56-226.8)	
Colonoscopy findings				
CRC	258 (6.9)	138 (8.3)	120 (5.8)	
Proximal colon	62 (24)	25 (18)	37 (31)	0.025
Distal colon	97 (38)	51 (36)	46 (38)	
Rectum	99 (38)	62 (45)	37 (31)	
Stage I-II (N0,M0)	148 (57)	81 (59)	67 (56)	0.749
Stage III (N1-2,M0)	69 (27)	35 (25)	34 (28)	
Stage IV (M1)	20 (7.8)	11 (8.0)	9 (7.5)	
Stage unknown	21 (8.1)	11 (8.0)	10 (8.3)	
Advanced adenomas	1122 (30)	586 (35)	536 (26)	
Distal colon	370 (33)	196 (33)	174 (32)	
Proximal colon	120 (11)	61 (10)	59 (11)	
Several locations	516 (46)	272 (46)	244 (46)	
Location not stated	116 (10)	57 (9.7)	59 (11)	
<5mm	73 (6.5)	32 (5.5)	41 (7.6)	
5-10mm	277 (25)	130 (22)	147 (27)	
>10mm	758 (68)	415 (71)	343 (64)	
Size not stated	14 (1.2)	9 (1.5)	5 (0.9)	
Non-advanced adenomas/polyps	1031 (27)	453 (27)	578 (28)	
Distal colon	364 (35)	169 (37)	195 (34)	
Proximal colon	137 (13)	56 (12)	81 (14)	
Several locations	170 (16)	70 (15)	100 (17)	
Location not stated	360 (35)	158 (35)	202 (35)	
<5mm	516 (50)	223 (49)	293 (51)	
5-10mm	349 (34)	141 (31)	208 (36)	
>10mm (tex HP)	138 (13)	76 (17)	62 (11)	
Size not stated	28 (2.7)	13 (2.9)	15 (2.6)	
Other sources of bleeding	570 (15)	217 (13)	353 (17)	
Normal investigation	774 (21)	276 (17)	498 (24)	<0.05

Advanced adenomas= high-risk adenomas requiring follow-up colonoscopy. Non-advanced adenomas/polyps= all other polyps/adenomas. Other sources of bleeding= for example hemorroids, diverticulas, inflammation, angiodysplasia.

In Table 10 the number of CRC, AA, NNS and PPV for men and women at cut-off levels 40, 60, 80 and ≥100µg/g is specified. PPV for AA was significantly higher in men than in women with the current strategy and with cut-off 80µg/g in both genders. PPV for CRC was similar at cut-off level 80µg/g in both genders. Of the 120 women with CRCs, 28 (23%) had FIT level of 40-79µg/g.

Table 10. Colonoscopies required, advanced findings and PPV at different FIT cut-off levels and gender among 74,117 men and 84,032 women that participated.

FIT cut off level	Colonoscopies, n (% of FIT participants)	CRC, n	AA, n	NNS (CRC), n (95% C.I.)	NNS (AA), n (95% C.I.)	PPV (CRC), % (95% C.I.)*	PPV (AA), % (95% C.I.)**
Men, FIT ≥80µg/g	1672 (2.3)	138	586	12.1 (10.4-14.4)	2.9 (2.7-3.1)	8.3 (6.9-9.6)	35.0 (32.8-37.3)
Men, FIT ≥ 100µg/g	1436 (1.9)	128	523	11.2 (9.6-13.4)	2.7 (2.6-2.9)	8.9 (7.4-10.4)	36.4 (33.9-38.9)
Women, FIT ≥40µg/g	2086 (2.5)	120	536	17.4 (14.8-21.0)	3.9 (3.6-4.2)	5.8 (4.8-6.8)	25.7 (23.8-27.6)
Women, FIT ≥60µg/g	1501 (1.8)	99	401	15.2 (12.7-18.7)	3.7 (3.5-4.1)	6.6 (5.3-7.9)	26.7 (24.5-29.0)
Women, FIT ≥80µg/g	1193 (1.4)	92	328	13.0 (10.8-16.1)	3.6 (3.3-4.0)	7.7 (6.2-9.2)	27.5 (25.0-30.0)
Women, FIT ≥ 100µg/g	987 (1.2)	86	273	11.5 (9.5-14.4)	3.6 (3.3-4.0)	8.7 (7.0-10.5)	27.7 (24.9-30.5)
Total	3758 (2.4)	258	1122	14.6 (13.0-16.5)	3.3 (3.2-3.5)	6.9 (6.1-7.7)	29.9 (28.4-31.3)

Colonoscopies= Complete colonoscopies in FIT positive participants at each FIT cut off level. AA= High-risk adenomas that required follow-up colonoscopy. Positive predictive value, number of CRC or AA at each FIT cut off level divided by the number of complete colonoscopies. NNS= Number needed to scope, number of colonoscopies required for detecting one CRC or AA. *) PPV for CRC is higher in men than in women with current strategy (p=0.003), and equal with cut off 80µg/g for both gender (p=0.648). **) PPV for AA is higher in men than in women with current strategy (p <0.05), and with cut off 80µg/g for both gender (p <0.05).

The screening costs of the study period with the current strategy are summarized in Table 11 with an estimation of costs at cut-off level of 80µg/g in both genders. Of the total running costs, the current strategy was 16% more expensive than the gender-equal strategy, corresponding to a 3% increment per detected CRC. A sensitivity analysis excluding the follow-up colonoscopies rendered similar results.

Table 11. Estimated costs for Stockholm-Gotland screening program with current strategy of cut-off of 80µg/g for men and 40µg/g for women, and expected costs with equal cut-off of 80µg/g.

Post	Current strategy, n	Cost current strategy, SEK	Cut-off 80µg/g for men and women, n	Cost cut-off 80 µg/g for men and women, SEK
Staff and administration	per two years	9,230,536	per two years	9,230,536
Invitation and FIT kit	229,944	6,360,251	229,944	6,360,251
Reminder	99,874	667,158	99,874	667,158
FIT analysis	158,149	12,474,793	158,149	12,474,793
Re-test and analysis	7753	826,005	7753	826,005
Re-test not analyzed	1274	35,239	1274	35,239
Index colonoscopies*	3758	24,472,096	2865	18,656,880
Follow-up colonoscopies	1122	7,306,464	914	5,951,968
Total running costs		52,142,006		44,972,294
CRC detected	258		230	
Cost per detected CRC		202,101		195,532

*) Index colonoscopies after exclusions

5.4 Paper IV

For the first screening round 214,356 were invited and 69% participated. In the cohort 257 SD-CRCs, 124 FIT ICs, 7 colonoscopy ICs, 3 ICs in individuals non-compliant to colonoscopy and 177 CRCs in non-participants were detected within 2 years. The IC rate was higher in men than in women (12.6 vs 6.0 per 10,000 negatives, p=0.00005) (Table 12).

Table 12. FIT results, SD-CRC and IC rates in different age and gender subgroups in the Stockholm-Gotland screening program 2015-2017.

Age at invitation and gender	FIT Participants	FIT negatives	FIT positives	Colono-scopies	SD-CRC	FIT IC	Colonoscopy IC	IC non-compliant to colonoscopy	IC rate, (95% CI)*	Test Sensitivity, (95% CI)**	IC incidence rate (95% CI)
Women <65	46,460	45,259	1201	1043	59	20	2	0	4.8 (2.8-6.7)	0.75 (0.65-0.84)	23.8 (15.7-36.1)
Women ≥65	31,680	30,669	1011	901	63	21	3	1	7.9 (4.8-11.0)	0.75 (0.66-0.84)	39.7 (26.8-58.7)
Men <65	42,159	41,113	1046	910	72	37	2	2	9.8 (6.8-12.8)	0.66 (0.57-0.75)	48.9 (36.0-66.4)
Men ≥65	26,679	25,880	799	667	63	46	0	0	17.4 (12.4-22.4)	0.58 (0.49-0.67)	86.8 (65.1-115.9)
All	146,978	142,921	4057	3521	257	124	7	3	9.2 (7.6-10.7)	0.68 (0.63-0.72)	45.8 (38.7-54.3)

FIT IC= IC after negative FIT. Colonoscopy IC= IC after negative screening colonoscopy. IC rate= number of IC per 10,000 FIT negatives or FIT positives with negative colonoscopy. Test sensitivity = SD/(SD+FIT IC). IC incidence rate= number of IC among FIT negatives or FIT positives with negative colonoscopy per 100,000 person-years of follow-up. *) p-value= 0.000053 for difference in IC rate between men and women. **) p-value= 0.019 for difference in test sensitivity between men and women ≥65.

The IC rate was higher in the second as compared to the first year after invitation in each subgroup (Table 13). The rate ratio of the IC incidence/EIR was 0.30-0.44 and non-significantly lower in the women as compared to the men in each age group.

Test sensitivity was higher in women than in men (0.75 vs 0.62, p-value 0.011), but equal had cut-off level been 80µg/g in both genders (0.56 vs 0.62, p-value 0.259). Moreover, test sensitivity was significantly higher in distal vs proximal CRC (0.75 vs 0.52).

In all the 568 CRCs including those in the non-participants, proximal localization was more common in women (42%) than in men (29%) (p-value 0.0030). In the SD-CRC the proportion of stage I&II (55.3%) and distally located CRC (74.7%) was higher than that of FIT IC and CRCs in non-participants.

Table 13. IC incidence rate in relation to experienced incident rate (EIR) and by year from invitation in the Stockholm-Gotland screening program 2015-2017.

Age at invitation and gender	Total IC, N	IC incidence rate (95% CI)	EIR	IC incidence rate/EIR (95% CI)	IC 0-12 months from invitation				IC 13-24 months from invitation			
					Observed ICs	IC rate (95% CI)	IC incidence rate (95% CI)	IC incidence rate/EIR (95% CI)	Observed ICs	IC rate (95% CI)	IC incidence rate (95% CI)	IC incidence rate/EIR (95% CI)
Women <65	22	23.8 (15.7-36.1)	78.3	0.30 (0.18-0.49)	7	3.02 (1.21-6.22)	30.2 (12.1-62.2)	0.39 (0.15-0.84)	15	6.47 (3.62-10.7)	64.7 (36.2-106.6)	0.83 (0.44-1.45)
Women ≥65	25	39.7 (26.8-58.7)	131.8	0.30 (0.19-0.46)	6	3.80 (1.39-8.26)	38.0 (13.9-82.6)	0.29 (0.10-0.64)	19	12.0 (7.24-18.8)	120.2 (72.4-187.7)	0.91 (0.53-1.48)
Men <65	41	48.9 (36.0-66.4)	123.8	0.39 (0.27-0.57)	11	5.23 (2.61-9.35)	52.3 (26.1-93.5)	0.42 (0.21-0.78)	30	14.3 (9.62-20.4)	142.6 (96.2-203.5)	1.15 (0.74-1.72)
Men ≥65	46	86.8 (65.1-115.9)	198.3	0.44 (0.31-0.61)	15	11.3 (6.31-18.6)	112.7 (63.1-185.9)	0.57 (0.31-0.96)	31	23.3 (15.8-33.1)	232.9 (158.3-330.6)	1.18 (0.78-1.72)

EIR= Experienced incidence rate; mean value in Stockholm-Gotland region for the years 1998-2007 in different age and gender groups per 100,000. IC incidence rate= number of IC per FIT negatives or FIT positives with negative colonoscopy per 100,000 person-years. IC rate= number of IC per 10,000 FIT negatives or FIT positives with negative colonoscopy.

6 DISCUSSION

6.1 General discussion

The main findings regarding the screening strategy in paper I and II are that one FIT sample at lower cut-off level performed better than two samples at a higher cut-off level in terms of CRC detection. Moreover, that sensitivity for AN at each of the cut-off levels did not increase with the second sample but decreased the specificity at the lower cut-off levels. This implies using a single sample and instead modify the cut-off level according to the available colonoscopy capacity. The finding is in line with the previously cited randomized trial on single vs multiple samples. Although participation did not decrease with two samples the authors recommended the single sample screening strategy because the cumulative AN detection was similar between strategies and the colonoscopy demand lower for the single sample strategy (103, 105). In paper II the sensitivity and specificity measures referred mainly to AAs because there was only one CRC detected (0.1%), hence the diagnostic yield was low but comparable to other colonoscopy screening cohorts (60, 108, 130). Moreover, 5-6% of participants in paper I and II took the two samples on the same date which may also have hampered the diagnostic yield and likewise points to using a single sample strategy.

Regarding the cut-off level, the rate of missed AN and CRC increased with the cut-off level, but more so for AN than for CRC. This is because CRC is a stronger source of bleeding than AA, so the cut-off level has a more pronounced effect of AA detection than it has for CRC, as seen in previous studies (107, 129).

Paper I and II assessed the FIT level in association to adenoma characteristics. Why is this important? AAs are precursors to CRC, and the adenomas removed in screening are the ones that could contribute to a decreased incidence of CRC eventually seen in screened populations (157, 159). FIT is not a perfect test for detecting CRC or adenomas. It detects blood in the stool; hence the lesion needs to be at least intermittingly bleeding to be detected by the test. If some lesions are less likely to bleed, they will be more frequently missed by FIT, and if these non-bleeding lesions are more frequent in certain population groups screening will be less beneficial for them. This is in particular important for SSAs -precursors to mainly proximal CRC for which FIT screening is less protective (115).

In paper I the Hemoglobin level of the first FIT sample was independently correlated to the adenoma size and not to gender, whereas in paper II FIT positivity was more often seen in adenomas with pedunculated shape, high-risk dysplasia and in men but not in those with large size. The cause of the results to differ between the papers are several and has been seen previously across many other studies cited in the literature review.

Firstly, these features are correlated to one another, i.e., a large adenoma is more likely to exhibit HGD and a pedunculated adenoma is more likely to be large. In some cohorts the size would be the independent characteristic significantly associated to FIT positivity, and in others the grade of dysplasia or a pedunculated shape. One must also remember that the size

was measured by the endoscopist in some studies (as in this thesis) and by the pathologist in others. Some studies have indicated that polyps shrink when put in formaldehyde which may influence the size measurement, and flat or sessile morphology is more likely to be overestimated by endoscopists (175). Moreover, the grade of dysplasia as part of the advanced and non-advanced features of adenomas may be judged differently by different pathologists, hence assessment by expert gastrointestinal pathologists at a tertiary hospital may differ from that of general pathologist in rural hospitals (176).

Secondly, in paper I the adenoma characteristics were assessed in relation to the median FIT level of the FIT positives' first sample and in paper II in relation to FIT positivity at a low cut-off level (10µg/g in any of the two samples). The association between e.g., size and degree of bleeding might be evident at higher cut-off levels. Moreover, some statistical strength was lost partly because paper II displayed half the number of participants with adenoma compared to paper I and partly because of the categorization of FIT levels into positive/negative.

Neither in paper I or paper II the adenoma localization was independently associated with FIT level or FIT positivity, and the results from previous studies have been conflicting (35, 96, 99). This could be related to the other adenoma characteristics: in some of the studies the proximal adenomas were larger than the distal, and distal adenomas are more often pedunculated, though some of the studies controlled for this.

FIT performance has been evaluated in several screening studies, but as pointed out previously the study population and the settings differ substantially. The FIT performance depends on the study population e.g., the age and gender distribution and the presence of other risk factors for adenoma or CRC formation. The predictive values are determined by the prevalence in the population, hence in a high-risk population the FIT performance is better than in a low- or average-risk population. In a study where the subjects display a high rate of proximal and flat adenomas (e.g., women) the performance could be worse than in that with a high rate of distal pedunculated adenomas (e.g., men). Indeed, the sensitivity was lower in paper II than in previously evaluated European cohorts.

Moreover, the settings vary between studies which may influence the results. Most of the randomized controlled studies and screening pilot studies were conducted in an academic setting at tertiary (university) hospitals with gastroenterologists performing the endoscopies and expert gastrointestinal pathologists evaluating the polyp and CRC specimens. This differs from that of a screening program or screening study in Sweden involving both small rural hospitals, university hospitals and nurses and surgeons performing some of the endoscopies. The participation of a randomized trial could also differ from that in a screening program, discussed in a later section.

Therefore, randomized trials or screening studies from other countries might not be generalizable and the evaluation of the Swedish screening setting is crucial and urgent before screening is being implemented nationally.

Gender-differences in screening were assessed in paper I and II, and a gender-specific screening strategy with lower cut-off levels in women was evaluated in paper III and IV.

With gender-uniform screening paper I and II demonstrated as expected a higher detection rate of AN in men and a higher PPV for AN in men at all the lower cut-offs. In paper II the sensitivity for AA was significantly higher in men than in women at the most sensitive screening strategy. However, the proportion of proximal lesions were similar in men and women in both the FIT positive and the colonoscopy cohorts and could not explain the gender differences in detection, but the association between FIT positivity in those with any adenoma and male gender could be related to more men taking NSAID and ASA medication.

In paper III and IV a gender-specific screening strategy rendered a high rate of normal colonoscopies in women as compared to men. As indicated in paper I, an equal AN detection rate could be accomplished with lowered cut-off levels in women at expense of a 30% increase in demand for colonoscopies. The corresponding estimates for CRC in paper I when raising the cut-off level from 40 to 80µg/g for the first sample were that CRC detection decreased from 19 to 17 CRCs (10.5%); 12.5% in women and 9% in men. However, in paper III almost a fourth of the female CRCs were found in the lowest FIT category with the gender-specific strategy, and because of this there was only a minor increase of cost per detected CRC. Meulen et al investigated multiple gender-specific strategies in a model-based study and concluded that gender-specific screening was not more cost-efficient, which could seem contradictory since cost-efficiency was higher in men. However, as FIT sensitivity is higher and AN more common in men there is a high efficiency for initiating screening in men but lower yield at subsequent rounds, whereas in women the lower prevalence gives a lower efficiency in initiating screening than in men but higher yield in subsequent rounds as compared to men, which evens out the differences (168).

Although the IC rate was significantly higher in men than in women with gender-specific screening in paper IV, the program missed CRCs in approximately the same rate as they are expected to appear in men and women. This implies that the IC incidence reflects the background CRC incidence: in age- or gender groups where the CRC incidence is lower, the IC incidence is also expected to be lower. However, there was a tendency towards a lower IC incidence rate relative to the background incidence in women compared to men, so larger studies on multiple screening rounds are warranted.

Applying a lower cut-off level in women yielded a higher test sensitivity in women as compared to men. The test sensitivity is the SD-CRC relative to the total number of CRCs, in contrast to the IC incidence relative to the CRC background incidence. In gender-uniform screening programs the test sensitivity has been lower, and the proportional IC incidence has been higher in women than in men (148). The test sensitivity is likely to decrease with multiple screening rounds as most of the prevalent CRCs are detected in the first round in the participants. The IC incidence is also likely to decrease with multiple screening rounds since more CRCs are detected over multiple rounds (147). Furthermore, the test sensitivity is

dependent on the participation rate since only participants contribute to SD-CRCs, hence age or gender groups with a higher participation rate affect the test sensitivity.

Stockholm-Gotland is the first screening program applying a gender-specific strategy, and the findings in relation to costs and the ICs have never been explored in an established program before. A similar strategy is being rolled out in Finland, and a pilot cohort study is requested in Scotland (127, 132). In the previous gFOBT program in Stockholm-Gotland, women were at disadvantage compared to men regarding program sensitivity. In the Finnish gFOBT randomized study there was no gain in CRC mortality between screened and unscreened and a higher CRC mortality rate in women than in men (177, 178). Moreover, in the 30-years of follow-up of the Minnesota trial the mortality reduction was non-significant in women for biennial screening (179). The most important finding from the ongoing Stockholm-Gotland screening program is that the test sensitivity and the IC rate have improved as compared to the gFOBT program and that the disadvantage for women seems to have been overcome regarding the studied outcomes even though the magnitude is uncertain, but future studies will have to address the results from multiple screening rounds in terms of CRC mortality.

6.2   Methodological considerations

All four papers in the thesis are population-based cohort studies, in paper I-II the study cohorts were derived from a randomized screening trial and in paper III-IV from two partly overlapping screening cohorts in the Stockholm-Gotland screening program. In terms of screening strategy, the exposure was the FIT screening in men and women, and the outcome was the colonoscopy findings or the screening cost. In terms of FIT accuracy, the exposure was the colonoscopy findings, and the outcome was the FIT level or FIT positivity (in men and women). There are different types of bias relevant in screening, as already discussed in the background: selection bias, information bias and lead time bias. A discussion of the impact of bias, confounding and the validity in relation to papers I-IV will follow.

Selection bias

Selection bias occurs when the risk of the outcome among exposed and non-exposed participants differs from that of the non-participants, so that the choice of participating or not is influenced by the risk. The choice to participate in a screening trial or in a screening program could be affected by e.g., family history of CRC leading to high-risk individuals choosing to participate relative to non-participants, or the opposite; a selection of health-conscious low-risk individuals. This would affect the generalizability of the study but not the internal validity (within the study cohort): It is not likely that high-risk individuals bleed more from their AN than average-risk individuals. Caution should always be taken when comparing FIT cohorts and colonoscopy cohorts due to the possibility of selection bias in colonoscopy cohorts with low participation rates selecting more healthy participants with fewer findings and hence lower precision and generalizability.

However, there are indications that blood in the stool is a marker for general inflammation and poor health and associated with all-cause mortality (180). If men and women display

different mortality risk, the FIT level could serve as a confounding factor when evaluating mortality.

Lead-time bias

Lead-time bias is as explained earlier an observed benefit in survival related to the early diagnosis in screening compared to a later-stage clinical diagnosis in non-screened. A lead time bias would appear in paper IV if survival was measured in screening participants as compared to non-participants, since most SD-CRC were early stage and the time elapsed from screening to the appearance of symptoms would have been included in the survival time. This is avoided when comparing disease mortality, which is however not the intended outcome measure for this thesis. A lead time bias could be introduced if the screening colonoscopy in the FIT positives were more delayed in those with higher FIT levels as compared to those with lower levels. This could occur theoretically in participants with high FIT levels and stricturing CRCs in which the colonoscopy had to be rescheduled due to insufficient bowel cleansing, thereby allowing the CRC to grow into more advanced stage. This is probably of limited importance since few tumors were stricturing. Another scenario would be participants with a higher age and hence higher FIT levels, who would, due to their age, have difficulties with bowel preparation, thus being rescheduled for colonoscopy at a later point during which the CRC was allowed to progress to more advanced stage. In paper III one of the exclusion criteria was >6 months elapsed from FIT sample to colonoscopy investigation to ensure the timely relation between colonoscopy finding and FIT level. Paper I and II was conducted in an age-homogeneous cohort.

Length-biased sampling

Length bias sampling has been acknowledged in e.g., breast cancer screening and implies that slow-growing cancers are more likely to be detected at screening than fast-growing cancers. Certain biological features, e.g., hormonal receptor status, of slow-growing breast cancers are associated with the response to treatment and hence to the prognosis, even if they are diagnosed in the same stage as a clinically detected cancer, which overestimates the benefit of screening on survival (181). For CRC, the natural course of disease might not be fully understood as there is a possibility of asymptomatic non-progressive cancers that are never diagnosed (182). In the four randomized controlled trials comparing gFOBT screening to no screening there was a higher rate of early-stage CRC (Dukes A) in the screening arms as compared to the control arms, hence the CRC mortality reduction with screening was attributable to the early detection (52). However, if a proportion of the CRCs in the control-arm were slow-growing and not even detected clinically, the randomization did not account for the length bias and the gain in CRC mortality with screening was overestimated. In the Taiwanese FIT screening program, the relative CRC mortality in screened vs. unscreened was 0.72 and 0.56 for proximal and distal CRC respectively, because proximal SD-CRCs were more often in a later stage. Moreover, the authors hypothesize that proximal CRCs could be more fast-growing and confer a worse prognosis than distal and are less likely to be diagnosed in FIT screening (183). Therefore, the tumor biology in screen-detected vs. non-

screening detected CRCs and length-biased sampling needs to be assessed in future evaluations of disease mortality of the Stockholm-Gotland program.

Information bias -misclassification bias

In cohort studies misclassification of either the exposure or outcome could occur. If the misclassification is dependent of the outcome or the exposure it is differential, otherwise it is non-differential. Usually in cohort studies, differential misclassification could arise from loss to follow-up e.g., if cancer registers are lacking or are incomplete or if individuals are not traceable. In Sweden every individual has a unique personal identification number, and the cancer register has almost 100% coverage. However, in a screening setting misclassification of adenomas could occur either at the colonoscopy (missed lesions) or at the pathological examination of polyps. All adenoma and CRC findings in the study register in paper I and II were verified against the pathology report, and in paper III and IV CRCs were identified in the SCRCR. A misclassification of adenomas as the exposure would dilute the association with FIT level e.g., SSA misclassified as being HP, small adenomas misclassified as large, or AA misclassified as non-AA. Most likely this would be non-differential and occur in all categories of comparison. The FIT level is machine read and as such more robust (compared to gFOBT that was read visually). However, when analyzing the data in the Stockholm-Gotland program 52 participants with FIT results exactly on the cut-off levels were discovered having had an erroneous negative result in their reply letter, thereby not being offered colonoscopy. On the other hand, the analyses were done as an intention-to-treat referring to the actual cut-off level and not the positive/negative test result, therefore being a non-differential misclassification of exposure weakening the association of the colonoscopy findings.

Confounding

Confounding occurs when a factor (measured or unmeasured in the study) is related to both the exposure and the outcome. For instance, if a study revealed that female gender was positively associated to CRC (sic), the age of the participants could be a confounder, and the results explained by the women being older than the men thereby having a higher risk of CRC, rather than by the gender itself. Confounding could be dealt with through stratifying, by adjusting for the confounder in a regression model or by randomization to control for unknown confounders.

In paper I the association of colonoscopy findings to FIT was adjusted for medication, gender, other findings, and BMI meaning that the association was explored holding these variables constant. BMI is a confounder to colonoscopy findings because obesity is associated both to CRC and adenoma formation and progression and to the colonoscopy quality i.e., caecal intubation, and hence could be related to the adenoma detection, in addition to being a risk factor for false positive FIT. The colonoscopy quality parameters were generally very high and not related to the FIT level, nor adjusted for. In paper II the association between FIT positivity and the adenoma characteristics was assessed in a

multivariable analysis, and instead a separate sensitivity analysis was used where those with adenoma and NSAID or ASA medication were excluded. In this analysis the association between FIT positivity and gender did not remain, suggesting that the higher chance of being FIT positive when having an adenoma was related to the medication and that male gender was a confounder. By stratifying the adenoma participants by gender, we could see that the number of adenomas differed in men and women, but this variable was not associated with FIT positivity. The size and localization distributions were equal in men and women and could neither explain the gender differences in FIT positivity.

In paper III a sensitivity analysis of the screening cost was conducted to see if the differences between screening strategies held true without regarding the follow-up colonoscopies in the polyp surveillance program -as these colonoscopies could be viewed as not directly related to the screening. The magnitude of the difference remained; hence the follow-up colonoscopies were not decisive.

In paper IV the estimates of the IC incidence rate in relation to CRC incidence was stratified in different age and gender groups. An alternative method would be to age- and gender standardize the rates and present a composite rate ratio, but the very point of the evaluation was to compare the rate ratios between strata in a gender-specific screening program.

In both paper I and II the invitations were sent to randomly selected 60-year-olds, but the randomization referred to screening mode (colonoscopy vs FIT vs non-screening) on the outcome disease mortality, and not to the outcomes relevant to this thesis. Nevertheless, the invited cohorts should represent the Swedish average-risk population of 60-year-olds.

Interactions

Interaction are circumstances that could affect the magnitude of the outcome estimates, such as endoscopists specialized in gastroenterology and expert gastrointestinal pathologists being more accurate in detecting and diagnosing SSA compared to others (184, 185). This expertise is more likely to be present in tertiary hospital settings and depending on the proportion of participants referred to such centers this would modify the estimates of association. We have not stratified the results in paper I-II by endoscopy and pathology center, so this effect is unknown. However, the participants in paper I-II are randomly invited to screening and should mimic the general population, hence overall be generalizable to the total population eligible to screening.

Random error

A random error refers to results obtained by chance and is reflected in the precision of the estimates. The p-value is used in hypothesis testing to describe the probability of getting the result, given that the null hypothesis is true and is usually set to a significance level of 0.05. This means that there is a 5% chance of getting this or more extreme results by chance, even though the null hypothesis is true. Falsely rejecting the hypothesis is called a type I error and failing to reject a false hypothesis is called a type II error. Another precision measure is the

CI, which is presented with the estimate and describes the range where the true population value is likely to be found, usually set to 95%. It is calculated from the standard error which depends on the number of observations in the data set. A larger study gives a smaller range of the CI and higher precision and vice versa. A 95% CI means that with 95% certainty the true value is found within this range. A 90% CI would generate a smaller range but a larger uncertainty.

In paper I-IV some of the subgroups were small, which made precision low. The number of CRCs in paper I-II was low and differences in PPV for CRC between men and women non-significant with overlapping 95% CI. In paper III it was not possible to analyze the potential associations of CRC stage and localization in women with different FIT categories.

In paper IV the number of CRC in the different age and gender groups were small and hence the 95% CI of the estimated IC incidence/EIR were wide and overlapping. There was a tendency towards lower rate ratio in women than in men, but precision was low. It is possible that with a larger study i.e., several screening rounds, the estimates would be more precise, and a different conclusion to be drawn. Putting all ages together would have made the subgroups larger but precluded conclusions of the effect of age.

External validity

The internal validity refers to the validity of the results in the study, the robustness of the estimates discussed above. The external validity is the extent to which the results could be generalized to the target population of the study. The external validity is the major strength of this thesis as paper I-IV are conducted in Swedish screening settings in which the results are supposed to apply. However, there are some concerns. Paper I-II are conducted in screening trial cohorts, and the matter of self-selection of people willing to volunteer in a screening trial -as discussed above, could affect the generalizability to a screening program. Furthermore, these studies were conducted in 60-year-olds which might explain the evidently differing estimates in detection as compared to paper III-IV. In paper III-IV the study population was derived from the Stockholm-Gotland region. Although this region consists of both rural and urban areas the distances to health service and university hospital expertise is smaller than in other countryside areas in Sweden e.g., in the north. Moreover, the estimates of the CRC incidence prior to screening in the region may not be generalizable to other regions in Sweden.

Strengths

The strengths of paper I-IV are as mentioned above the generalizability to Swedish screening settings and in paper III-IV the first evaluation of an ongoing population-based gender-specific screening program. Moreover, the ability of confirming all adenoma and CRC findings in the pathology reports or in the SCRCR reduces the risk of misclassification of findings. This increases the internal and external validity of the thesis.

Implications

Colorectal cancer screening is now being implemented at a national level in Sweden and the evaluation of the Stockholm-Gotland screening program as well as results from a national screening study are relevant when determining the national screening strategy. The colonoscopy resources and the structure of the target population differ between regions and must guide this decision, hopefully with the aid of the estimated colonoscopy requirements at different cut-off levels presented in this thesis.

However, neither a gender-uniform nor a gender-specific FIT screening strategy have been evaluated with regards to CRC mortality, which is the ultimate endpoint of the screening. Nevertheless, given the results from previous gFOBT trials, it does not seem ethically legitimate to await the results from the ongoing FIT studies on mortality before implementing national screening, since FIT has shown a higher sensitivity for AN than gFOBT.

Previous studies as well as this thesis have indicated that a gender-uniform screening strategy is less beneficial in women as compared to men regarding AN detection. Furthermore, the gender-specific screening strategy with lower cut-off levels in women could balance some of the differences in men and women at a minor increase of the screening costs, which could justify this strategy. The IC rate was higher in men than in women with a gender-specific screening that applies lower cut-off levels in women, but taking the background incidence into consideration i.e., the higher incidence of CRC in men, the IC incidence appears more equal, although larger studies are required to fully review the consequences.

7 CONCLUSIONS

• In terms of CRC detection, a single FIT sample at lower cut-off level is advocated instead of two samples at a higher threshold. The cut-off level could be adjusted to the available colonoscopy resources.

• In FIT screening of a Swedish average-risk population individuals with CRC and AA displayed significantly higher FIT levels than in those with non-advanced or normal findings. FIT more often detected high-risk adenomas that requires follow-up colonoscopy, but sensitivity for AA was lower in women than in men.

• Lower cut-off levels in women yielded a high rate of normal colonoscopies in women, but as almost a fourth of CRCs in women were found in the lowest FIT category there was only a minor increment in screening cost per detected CRC.

• Lower cut-off levels in women rendered a higher IC rate and a lower test sensitivity in men as compared to women. However, larger studies of this gender-specific screening strategy are needed to fully review the ICs relative to the background incidence in men and women.

8 POINTS OF PERSPECTIVE

Results from the ongoing randomized trials of FIT screening as compared to colonoscopy on the disease mortality is expected soon (November 2021) for COLONPREV, and in 2028 and 2034 for CONFIRM and SCREESCO respectively. However, regarding colonoscopy screening, the published results and results from other colonoscopy trials and screening programs suffers from low participation. Screening is only defendable if it is broadly accepted in the population and a lower CRC mortality in the population can only be achieved if people choose to participate.

Furthermore, the efficacy in randomized trials must be explored in an actual screening program, and the effect on CRC mortality in the previous Stockholm-Gotland gFOBT program as well as in the ongoing gender-specific FIT program needs to be assessed. Moreover, investigating the cost-efficiency of gender-specific screening programs await, and simulation models must be based on real data from the screening program (186). A continued high participation rate is crucial for mortality reduction, and many evaluations of efforts to increase participation are made and have been made.

A positive "side-effect" of organized screening and allocation of colonoscopy resources is the attention to colonoscopy quality and the development of a national colonoscopy quality register that will serve as a base for future research and improvements in education and quality.

Of other screening modalities, DNA fecal tests are under development and alone or in combination with FIT could serve as screening tool but has the disadvantage of lower specificity and higher cost. If considerably cheaper in the future this could be an attractive alternative.

9 ACKNOWLEDGEMENTS

Ingrid & Bodil, älskade barn, för att ni får mig att tänka på något annat

10 REFERENCES

1.   Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: a cancer journal for clinicians. 2021;71(3):209-49.

2.   National Board of Helth and Welfare cancer statistics 2018 available from http://www.socialstyrelsen.se/statistik/statistikdatabas/cancer Accessed: November 2021

3.   Report from the Swedish Colorectal Cancer Register 2020 available at https://cancercentrum.se/globalassets/cancerdiagnoser/tjock--och-andtarm-anal/kvalitetsregister/tjock--och-andtarm-2021/kolonrapport.pdf Accessed: November 2021

4.   Brenner H, Kloor M, Pox CP. Colorectal cancer. Lancet. 2014;383(9927):1490-502.

5.   UJCC TNM Classification of Malignant Tumors, 8:th Edition, p74-76. Brierley, JD et al. Wiley Blackwell 2017.

6.   Kuhry E, Schwenk WF, Gaupset R, Romild U, Bonjer HJ. Long-term results of laparoscopic colorectal cancer resection. The Cochrane database of systematic reviews. 2008(2):CD003432.

7.   Argiles G, Tabernero J, Labianca R, Hochhauser D, Salazar R, Iveson T, et al. Localised colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2020;31(10):1291-305.

8.   Glimelius B, Tiret E, Cervantes A, Arnold D, Group EGW. Rectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2013;24 Suppl 6:vi81-8.

9.   Biller LH, Schrag D. Diagnosis and Treatment of Metastatic Colorectal Cancer: A Review. Jama. 2021;325(7):669-85.

10.   Nagtegaal ID, Odze RD, Klimstra D, Paradis V, Rugge M, Schirmacher P, et al. The 2019 WHO classification of tumours of the digestive system. Histopathology. 2020;76(2):182-8.

11.   Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, et al. Genetic alterations during colorectal-tumor development. The New England journal of medicine. 1988;319(9):525-32.

12.   Muto T, Bussey HJ, Morson BC. The evolution of cancer of the colon and rectum. Cancer. 1975;36(6):2251-70.

13.	Worthley DL, Whitehall VL, Spring KJ, Leggett BA. Colorectal carcinogenesis: road maps to cancer. World journal of gastroenterology : WJG. 2007;13(28):3784-91.

14.	De Palma FDE, D'Argenio V, Pol J, Kroemer G, Maiuri MC, Salvatore F. The Molecular Hallmarks of the Serrated Pathway in Colorectal Cancer. Cancers (Basel). 2019;11(7).

15.	Boland CR, Goel A. Microsatellite instability in colorectal cancer. Gastroenterology. 2010;138(6):2073-87 e3.

16.	Ward R, Meagher A, Tomlinson I, O'Connor T, Norrie M, Wu R, et al. Microsatellite instability and the clinicopathological features of sporadic colorectal cancer. Gut. 2001;48(6):821-9.

17.	Suzuki H, Tokino T, Shinomura Y, Imai K, Toyota M. DNA methylation and cancer pathways in gastrointestinal tumors. Pharmacogenomics. 2008;9(12):1917-28.

18.	Thorlacius H, Takeuchi Y, Kanesaka T, Ljungberg O, Uedo N, Toth E. Serrated polyps - a concealed but prevalent precursor of colorectal cancer. Scandinavian journal of gastroenterology. 2017;52(6-7):654-61.

19.	Crockett SD, Nagtegaal ID. Terminology, Molecular Features, Epidemiology, and Management of Serrated Colorectal Neoplasia. Gastroenterology. 2019;157(4):949-66 e4.

20.	Eide TJ. Risk of colorectal cancer in adenoma-bearing individuals within a defined population. International journal of cancer Journal international du cancer. 1986;38(2):173-6.

21.	Martinez ME, Baron JA, Lieberman DA, Schatzkin A, Lanza E, Winawer SJ, et al. A pooled analysis of advanced colorectal neoplasia diagnoses after colonoscopic polypectomy. Gastroenterology. 2009;136(3):832-41.

22.	Stryker SJ, Wolff BG, Culp CE, Libbe SD, Ilstrup DM, MacCarty RL. Natural history of untreated colonic polyps. Gastroenterology. 1987;93(5):1009-13.

23.	Atkin WS, Valori R, Kuipers EJ, Hoff G, Senore C, Segnan N, et al. European guidelines for quality assurance in colorectal cancer screening and diagnosis. First Edition--Colonoscopic surveillance following adenoma removal. Endoscopy. 2012;44 Suppl 3:SE151-63.

24.	Kahi CJ, Hewett DG, Norton DL, Eckert GJ, Rex DK. Prevalence and variable detection of proximal colon serrated polyps during screening colonoscopy. Clinical gastroenterology and hepatology: the official clinical practice journal of the American Gastroenterological Association. 2011;9(1):42-6.

25.   JEG IJ, Bevan R, Senore C, Kaminski MF, Kuipers EJ, Mroz A, et al. Detection rate of serrated polyps and serrated polyposis syndrome in colorectal cancer screening cohorts: a European overview. Gut. 2017;66(7):1225-32.

26.   Thorlacius H, Bjork J, Ost A, Toth E. [Endoscopic surveillance after colorectal polypectomy]. Lakartidningen. 2017;114.

27.   Young GP, Symonds EL, Allison JE, Cole SR, Fraser CG, Halloran SP, et al. Advances in Fecal Occult Blood Tests: the FIT revolution. Digestive diseases and sciences. 2015;60(3):609-22.

28.   Tinmouth J, Lansdorp-Vogelaar I, Allison JE. Faecal immunochemical tests versus guaiac faecal occult blood tests: what clinicians and colorectal cancer screening programme organisers need to know. Gut. 2015;64(8):1327-37.

29.   Halloran SP, Launoy G, Zappa M, International Agency for Research on C. European guidelines for quality assurance in colorectal cancer screening and diagnosis. First Edition--Faecal occult blood testing. Endoscopy. 2012;44 Suppl 3:SE65-87.

30.   Chiang TH, Chuang SL, Chen SL, Chiu HM, Yen AM, Chiu SY, et al. Difference in performance of fecal immunochemical tests with the same hemoglobin cutoff concentration in a nationwide colorectal cancer screening program. Gastroenterology. 2014;147(6):1317-26.

31.   Lee JK, Liles EG, Bent S, Levin TR, Corley DA. Accuracy of fecal immunochemical tests for colorectal cancer: systematic review and meta-analysis. Annals of internal medicine. 2014;160(3):171.

32.   Imperiale TF, Gruber RN, Stump TE, Emmett TW, Monahan PO. Performance Characteristics of Fecal Immunochemical Tests for Colorectal Cancer and Advanced Adenomatous Polyps: A Systematic Review and Meta-analysis. Annals of internal medicine. 2019;170(5):319-29.

33.   Graser A, Stieber P, Nagel D, Schafer C, Horst D, Becker CR, et al. Comparison of CT colonography, colonoscopy, sigmoidoscopy and faecal occult blood tests for the detection of advanced adenoma in an average risk population. Gut. 2009;58(2):241-8.

34.   Brenner H, Tao S. Superior diagnostic performance of faecal immunochemical tests for haemoglobin in a head-to-head comparison with guaiac based faecal occult blood test among 2235 participants of screening colonoscopy. European journal of cancer. 2013;49(14):3049-54.

35.   Park DI, Ryu S, Kim YH, Lee SH, Lee CK, Eun CS, et al. Comparison of guaiac-based and quantitative immunochemical fecal occult blood testing in a population at average risk undergoing colorectal cancer screening. The American journal of gastroenterology. 2010;105(9):2017-25.

36.   Wong CK, Fedorak RN, Prosser CI, Stewart ME, van Zanten SV, Sadowski DC. The sensitivity and specificity of guaiac and immunochemical fecal occult blood tests for the detection of advanced colonic adenomas and cancer. International journal of colorectal disease. 2012;27(12):1657-64.

37.   Valori R, Rey JF, Atkin WS, Bretthauer M, Senore C, Hoff G, et al. European guidelines for quality assurance in colorectal cancer screening and diagnosis. First Edition--Quality assurance in endoscopy in colorectal cancer screening and diagnosis. Endoscopy. 2012;44 Suppl 3:SE88-105.

38.   Kothari ST, Huang RJ, Shaukat A, Agrawal D, Buxbaum JL, Abbas Fehmi SM, et al. ASGE review of adverse events in colonoscopy. Gastrointestinal endoscopy. 2019;90(6):863-76 e33.

39.   Pox CP, Schmiegel W. Role of CT colonography in colorectal cancer screening: risks and benefits. Gut. 2010;59(5):692-700.

40.   Stoop EM, de Haan MC, de Wijkerslooth TR, Bossuyt PM, van Ballegooijen M, Nio CY, et al. Participation and yield of colonoscopy versus non-cathartic CT colonography in population-based screening for colorectal cancer: a randomised controlled trial. The Lancet Oncology. 2012;13(1):55-64.

41.   Levine MS, Yee J. History, evolution, and current status of radiologic imaging tests for colorectal cancer screening. Radiology. 2014;273(2 Suppl):S160-80.

42.   Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP, et al. Multitarget stool DNA testing for colorectal-cancer screening. The New England journal of medicine. 2014;370(14):1287-97.

43.   Bosch LJW, Melotte V, Mongera S, Daenen KLJ, Coupe VMH, van Turenhout ST, et al. Multitarget Stool DNA Test Performance in an Average-Risk Colorectal Cancer Screening Population. The American journal of gastroenterology. 2019;114(12):1909-18.

44.   Naber SK, Knudsen AB, Zauber AG, Rutter CM, Fischer SE, Pabiniak CJ, et al. Cost-effectiveness of a multitarget stool DNA test for colorectal cancer screening of Medicare beneficiaries. PloS one. 2019;14(9):e0220234.

45.   Wilson JMG, Jungner G. Principles and practice of screening for disease. Geneva, Switzerland: WHO; 1968: Available at: http://apps.who.int/iris/bitstream/handle/10665/37650/WHO_PHP_34.pdf Accessed: February 2022

46.   European Colorectal Cancer Screening Guidelines Working G, von Karsa L, Patnick J, Segnan N, Atkin W, Halloran S, et al. European guidelines for quality assurance in colorectal cancer screening and diagnosis: overview and introduction to the full supplement publication. Endoscopy. 2013;45(1):51-9.

47.   National Board of Health and Welfare, statistics on causes of death in Sweden 2019, available at https://sdb.socialstyrelsen.se/if_dor/val.aspx Accessed: November 2021

48.   Kronborg O, Fenger C, Olsen J, Jorgensen OD, Sondergaard O. Randomised study of screening for colorectal cancer with faecal-occult-blood test. Lancet. 1996;348(9040):1467-71.

49.   Mandel JS, Bond JH, Church TR, Snover DC, Bradley GM, Schuman LM, et al. Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota Colon Cancer Control Study. The New England journal of medicine. 1993;328(19):1365-71.

50.   Hardcastle JD, Chamberlain JO, Robinson MH, Moss SM, Amar SS, Balfour TW, et al. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. Lancet. 1996;348(9040):1472-7.

51.   Lindholm E, Brevinge H, Haglind E. Survival benefit in a randomized clinical trial of faecal occult blood screening for colorectal cancer. The British journal of surgery. 2008;95(8):1029-36.

52.   Hewitson P, Glasziou P, Watson E, Towler B, Irwig L. Cochrane systematic review of colorectal cancer screening using the fecal occult blood test (hemoccult): an update. The American journal of gastroenterology. 2008;103(6):1541-9.

53.   Atkin WS, Edwards R, Kralj-Hans I, Wooldrage K, Hart AR, Northover JM, et al. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. Lancet. 2010;375(9726):1624-33.

54.   Schoen RE, Pinsky PF, Weissfeld JL, Yokochi LA, Church T, Laiyemo AO, et al. Colorectal-cancer incidence and mortality with screening flexible sigmoidoscopy. The New England journal of medicine. 2012;366(25):2345-57.

55.   Segnan N, Armaroli P, Bonelli L, Risio M, Sciallero S, Zappa M, et al. Once-only sigmoidoscopy in colorectal cancer screening: follow-up findings of the Italian Randomized Controlled Trial--SCORE. Journal of the National Cancer Institute. 2011;103(17):1310-22.

56.   Thiis-Evensen E, Hoff GS, Sauar J, Langmark F, Majak BM, Vatn MH. Population-based surveillance by colonoscopy: effect on the incidence of colorectal cancer. Telemark Polyp Study I. Scandinavian journal of gastroenterology. 1999;34(4):414-20.

57.   Holme O, Loberg M, Kalager M, Bretthauer M, Hernan MA, Aas E, et al. Effect of flexible sigmoidoscopy screening on colorectal cancer incidence and mortality: a randomized clinical trial. Jama. 2014;312(6):606-15.

58.   Holme O, Bretthauer M, Fretheim A, Odgaard-Jensen J, Hoff G. Flexible sigmoidoscopy versus faecal occult blood testing for colorectal cancer screening in

asymptomatic individuals. The Cochrane database of systematic reviews. 2013(9):CD009259.

59.   Hoff G, Grotmol T, Skovlund E, Bretthauer M, Norwegian Colorectal Cancer Prevention Study G. Risk of colorectal cancer seven years after flexible sigmoidoscopy screening: randomised controlled trial. Bmj. 2009;338:b1846.

60.   Bretthauer M, Kaminski MF, Loberg M, Zauber AG, Regula J, Kuipers EJ, et al. Population-Based Colonoscopy Screening for Colorectal Cancer: A Randomized Clinical Trial. JAMA internal medicine. 2016;176(7):894-902.

61.   Quintero E, Castells A, Bujanda L, Cubiella J, Salas D, Lanas A, et al. Colonoscopy versus fecal immunochemical testing in colorectal-cancer screening. The New England journal of medicine. 2012;366(8):697-706.

62.   Dominitz JA, Robertson DJ, Ahnen DJ, Allison JE, Antonelli M, Boardman KD, et al. Colonoscopy vs. Fecal Immunochemical Test in Reducing Mortality From Colorectal Cancer (CONFIRM): Rationale for Study Design. The American journal of gastroenterology. 2017;112(11):1736-46.

63.   The Screening of Swedish Colons (SCREESCO) study registred at https://clinicaltrials.gov/ct2/show/NCT02078804 Accessed: November 2021

64.   Randel KR, Schult AL, Botteri E, Hoff G, Bretthauer M, Ursin G, et al. Colorectal Cancer Screening With Repeated Fecal Immunochemical Test Versus Sigmoidoscopy: Baseline Results From a Randomized Trial. Gastroenterology. 2021;160(4):1085-96 e5.

65.   Force USPST, Davidson KW, Barry MJ, Mangione CM, Cabana M, Caughey AB, et al. Screening for Colorectal Cancer: US Preventive Services Task Force Recommendation Statement. Jama. 2021;325(19):1965-77.

66.   Wolf AMD, Fontham ETH, Church TR, Flowers CR, Guerra CE, LaMonte SJ, et al. Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society. CA: a cancer journal for clinicians. 2018;68(4):250-81.

67.   Altobelli E, Lattanzi A, Paduano R, Varassi G, di Orio F. Colorectal cancer prevention in Europe: burden of disease and status of screening programs. Preventive medicine. 2014;62:132-41.

68.   Schreuders EH, Ruco A, Rabeneck L, Schoen RE, Sung JJ, Young GP, et al. Colorectal cancer screening: a global overview of existing programmes. Gut. 2015;64(10):1637-49.

69.   Moss S, Ancelle-Park R, Brenner H, International Agency for Research on C. European guidelines for quality assurance in colorectal cancer screening and diagnosis. First

Edition--Evaluation and interpretation of screening outcomes. Endoscopy. 2012;44 Suppl 3:SE49-64.

70.   Kalager M, Wieszczy P, Lansdorp-Vogelaar I, Corley DA, Bretthauer M, Kaminski MF. Overdiagnosis in Colorectal Cancer Screening: Time to Acknowledge a Blind Spot. Gastroenterology. 2018;155(3):592-5.

71.   Carter JL, Coletti RJ, Harris RP. Quantifying and monitoring overdiagnosis in cancer screening: a systematic review of methods. Bmj. 2015;350:g7773.

72.   Wieszczy P, Kaminski MF, Loberg M, Bugajski M, Bretthauer M, Kalager M. Estimation of overdiagnosis in colorectal cancer screening with sigmoidoscopy and faecal occult blood testing: comparison of simulation models. BMJ open. 2021;11(4):e042158.

73.   Sanduleanu S, le Clercq CM, Dekker E, Meijer GA, Rabeneck L, Rutter MD, et al. Definition and taxonomy of interval colorectal cancers: a proposal for standardising nomenclature. Gut. 2015;64(8):1257-67.

74.   Steel MJ, Bukhari H, Gentile L, Telford J, Schaeffer DF. Colorectal adenocarcinomas diagnosed following a negative faecal immunochemical test show high-risk pathological features in a colon screening programme. Histopathology. 2021;78(5):710-6.

75.   Rothman KJ Modern epidemiology 3rd ed. Chapter 32. Philadelphia: Lippincott Williams & Wilkins, 2008

76.   The Swedish Agency for Health Technology Assessment and Assessment of Social Services (SBU), statistical methods manual, available at https://www.sbu.se/globalassets/ebm/metodbok/sbushandbok_bilaga10.pdf Accessed: November 2021

77.   Rothman KJ Modern epidemiology 3rd ed. Chapter 9. Philadelphia: Lippincott Williams & Wilkins, 2008

78.   Pinsky PF, Miller A, Kramer BS, Church T, Reding D, Prorok P, et al. Evidence of a healthy volunteer effect in the prostate, lung, colorectal, and ovarian cancer screening trial. American journal of epidemiology. 2007;165(8):874-81.

79.   Hakama M, Auvinen A, Day NE, Miller AB. Sensitivity in cancer screening. Journal of medical screening. 2007;14(4):174-7.

80.   Bretthauer M, Kalager M. Principles, effectiveness and caveats in screening for cancer. The British journal of surgery. 2013;100(1):55-65.

81.   Singal AG, Higgins PD, Waljee AK. A primer on effectiveness and efficacy trials. Clin Transl Gastroenterol. 2014;5:e45.

82.   Bulliard JL, Garcia M, Blom J, Senore C, Mai V, Klabunde C. Sorting out measures and definitions of screening participation to improve comparability: the example of colorectal cancer. European journal of cancer. 2014;50(2):434-46.

83.   Ness RM, Holmes AM, Klein R, Dittus R. Utility valuations for outcome states of colorectal cancer. The American journal of gastroenterology. 1999;94(6):1650-7.

84.   Drummond MF; Methods for the Economic Evaluation of Health Care Programs 4th ed. Chapter 5.4.3. Oxford: Oxford University Press; 2015

85.   Burstrom K, Sun S, Gerdtham UG, Henriksson M, Johannesson M, Levin LA, et al. Swedish experience-based value sets for EQ-5D health states. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2014;23(2):431-42.

86.   Drummond MF; Methods for the Economic Evaluation of Health Care Programs 4th ed. Chapter 4.2.1. Oxford: Oxford University Press; 2015

87.   National Board of Health and Welfare available from https://www.sbu.se/globalassets/ebm/metodbok/sbushandbok_kapitel11.pdf Accessed: November 2021

88.   Appleby J, Devlin N, Parkin D. NICE's cost effectiveness threshold. Bmj. 2007;335(7616):358-9.

89.   Ladabaum U. Cost-Effectiveness of Current Colorectal Cancer Screening Tests. Gastrointestinal endoscopy clinics of North America. 2020;30(3):479-97.

90.   Nimdet K, Chaiyakunapruk N, Vichansavakul K, Ngorsuraches S. A systematic review of studies eliciting willingness-to-pay per quality-adjusted life year: does it justify CE threshold? PloS one. 2015;10(4):e0122760.

91.   Drummond MF; Methods for the Economic Evaluation of Health Care Programs 4th ed. Chapter 9.4.5. Oxford: Oxford University Press; 2015

92.   Steele RJ, Rey JF, Lambert R, International Agency for Research on C. European guidelines for quality assurance in colorectal cancer screening and diagnosis. First Edition--Professional requirements and training. Endoscopy. 2012;44 Suppl 3:SE106-15.

93.   Wong MC, Ching JY, Chan VC, Lam TY, Shum JP, Luk AK, et al. Diagnostic Accuracy of a Qualitative Fecal Immunochemical Test Varies With Location of Neoplasia But Not Number of Specimens. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association. 2015;13(8):1472-9.

94.   Morikawa T, Kato J, Yamaji Y, Wada R, Mitsushima T, Shiratori Y. A comparison of the immunochemical fecal occult blood test and total colonoscopy in the asymptomatic population. Gastroenterology. 2005;129(2):422-8.

95.	Sohn DK, Jeong SY, Choi HS, Lim SB, Huh JM, Kim DH, et al. Single immunochemical fecal occult blood test for detection of colorectal neoplasia. Cancer research and treatment : official journal of Korean Cancer Association. 2005;37(1):20-3.

96.	Chiu HM, Lee YC, Tu CH, Chen CC, Tseng PH, Liang JT, et al. Association between early stage colon neoplasms and false-negative results from the fecal immunochemical test. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association. 2013;11(7):832-8 e1-2.

97.	Chiang TH, Lee YC, Tu CH, Chiu HM, Wu MS. Performance of the immunochemical fecal occult blood test in predicting lesions in the lower gastrointestinal tract. CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne. 2011;183(13):1474-81.

98.	Siripongpreeda B, Mahidol C, Dusitanond N, Sriprayoon T, Muyphuag B, Sricharunrat T, et al. High prevalence of advanced colorectal neoplasia in the Thai population: a prospective screening colonoscopy of 1,404 cases. BMC gastroenterology. 2016;16:101.

99.	de Wijkerslooth TR, Stoop EM, Bossuyt PM, Meijer GA, van Ballegooijen M, van Roon AH, et al. Immunochemical fecal occult blood testing is equally sensitive for proximal and distal advanced neoplasia. The American journal of gastroenterology. 2012;107(10):1570-8.

100.	Yuan SY, Wu W, Fu J, Lang YX, Li JC, Guo Y, et al. Quantitative immunochemical fecal occult blood test for neoplasia in colon cancer screening. Journal of digestive diseases. 2019;20(2):78-82.

101.	Rozen P, Levi Z, Hazazi R, Waked A, Vilkin A, Maoz E, et al. Identification of colorectal adenomas by a quantitative immunochemical faecal occult blood screening test depends on adenoma characteristics, development threshold used and number of tests performed. Alimentary pharmacology & therapeutics. 2009;29(8):906-17.

102.	Levi Z, Rozen P, Hazazi R, Vilkin A, Waked A, Maoz E, et al. A quantitative immunochemical fecal occult blood test for colorectal neoplasia. Annals of internal medicine. 2007;146(4):244-55.

103.	van Roon AH, Wilschut JA, Hol L, van Ballegooijen M, Reijerink JC, t Mannetje H, et al. Diagnostic yield improves with collection of 2 samples in fecal immunochemical test screening without affecting attendance. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association. 2011;9(4):333-9.

104.	Kapidzic A, van Roon AH, van Leerdam ME, van Vuuren AJ, van Ballegooijen M, Lansdorp-Vogelaar I, et al. Attendance and diagnostic yield of repeated two-sample faecal immunochemical test screening for colorectal cancer. Gut. 2017;66(1):118-23.

105.   Schreuders EH, Grobbee EJ, Nieuwenburg SAV, Kapidzic A, van Roon AHC, van Vuuren AJ, et al. Multiple rounds of one sample versus two sample faecal immunochemical test-based colorectal cancer screening: a population-based study. The lancet Gastroenterology & hepatology. 2019;4(8):622-31.

106.   Guittet L, Bouvier V, Mariotte N, Vallee JP, Levillain R, Tichet J, et al. Performance of immunochemical faecal occult blood test in colorectal cancer screening in average-risk population according to positivity threshold and number of samples. International journal of cancer Journal international du cancer. 2009;125(5):1127-33.

107.   Grazzini G, Visioli CB, Zorzi M, Ciatto S, Banovich F, Bonanomi AG, et al. Immunochemical faecal occult blood test: number of samples and positivity cutoff. What is the best strategy for colorectal cancer screening? British journal of cancer. 2009;100(2):259-65.

108.   Hernandez V, Cubiella J, Gonzalez-Mao MC, Iglesias F, Rivera C, Iglesias MB, et al. Fecal immunochemical test accuracy in average-risk colorectal cancer screening. World journal of gastroenterology: WJG. 2014;20(4):1038-47.

109.   Morikawa T, Kato J, Yamaji Y, Wada R, Mitsushima T, Sakaguchi K, et al. Sensitivity of immunochemical fecal occult blood test to small colorectal adenomas. The American journal of gastroenterology. 2007;102(10):2259-64.

110.   Haug U, Kuntz KM, Knudsen AB, Hundt S, Brenner H. Sensitivity of immunochemical faecal occult blood testing for detecting left- vs right-sided colorectal neoplasia. British journal of cancer. 2011;104(11):1779-85.

111.   Digby J, Fraser CG, Carey FA, McDonald PJ, Strachan JA, Diament RH, et al. Faecal haemoglobin concentration is related to severity of colorectal neoplasia. Journal of clinical pathology. 2013;66(5):415-9.

112.   Ciatto S, Martinelli F, Castiglione G, Mantellini P, Rubeca T, Grazzini G, et al. Association of FOBT-assessed faecal Hb content with colonic lesions detected in the Florence screening programme. British journal of cancer. 2007;96(2):218-21.

113.   Cubiella J, Castro I, Hernandez V, Gonzalez-Mao C, Rivera C, Iglesias F, et al. Characteristics of adenomas detected by fecal immunochemical test in colorectal cancer screening. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2014;23(9):1884-92.

114.   van Doorn SC, Stegeman I, Stroobants AK, Mundt MW, de Wijkerslooth TR, Fockens P, et al. Fecal immunochemical testing results and characteristics of colonic lesions. Endoscopy. 2015;47(11):1011-7.

115.   Chang LC, Shun CT, Hsu WF, Tu CH, Tsai PY, Lin BR, et al. Fecal Immunochemical Test Detects Sessile Serrated Adenomas and Polyps With a Low Level of Sensitivity. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association. 2017;15(6):872-9 e1.

116.   Haug U, Knudsen AB, Brenner H, Kuntz KM. Is fecal occult blood testing more sensitive for left- versus right-sided colorectal neoplasia? A systematic literature review. Expert Rev Mol Diagn. 2011;11(6):605-16.

117.   Hirai HW, Tsoi KK, Chan JY, Wong SH, Ching JY, Wong MC, et al. Systematic review with meta-analysis: faecal occult blood tests show lower colorectal cancer detection rates in the proximal colon in colonoscopy-verified diagnostic studies. Alimentary pharmacology & therapeutics. 2016;43(7):755-64.

118.   Lu M, Luo X, Li N, Chen H, Dai M. Diagnostic Accuracy Of Fecal Occult Blood Tests For Detecting Proximal Versus Distal Colorectal Neoplasia: A Systematic Review And Meta-Analysis. Clinical epidemiology. 2019;11:943-54.

119.   Niedermaier T, Balavarca Y, Brenner H. Stage-Specific Sensitivity of Fecal Immunochemical Tests for Detecting Colorectal Cancer: Systematic Review and Meta-Analysis. The American journal of gastroenterology. 2020;115(1):56-69.

120.   Nguyen SP, Bent S, Chen YH, Terdiman JP. Gender as a risk factor for advanced neoplasia and colorectal cancer: a systematic review and meta-analysis. Clinical gastroenterology and hepatology: the official clinical practice journal of the American Gastroenterological Association. 2009;7(6):676-81 e1-3.

121.   Ferlitsch M, Reinhart K, Pramhas S, Wiener C, Gal O, Bannert C, et al. Sex-specific prevalence of adenomas, advanced adenomas, and colorectal cancer in individuals undergoing screening colonoscopy. Jama. 2011;306(12):1352-8.

122.   Brenner H, Hoffmeister M, Arndt V, Haug U. Gender differences in colorectal cancer: implications for age at initiation of screening. British journal of cancer. 2007;96(5):828-31.

123.   Schoenfeld P, Cash B, Flood A, Dobhan R, Eastone J, Coyle W, et al. Colonoscopic screening of average-risk women for colorectal neoplasia. The New England journal of medicine. 2005;352(20):2061-8.

124.   Grobbee EJ, Wieten E, Hansen BE, Stoop EM, de Wijkerslooth TR, Lansdorp-Vogelaar I, et al. Fecal immunochemical test-based colorectal cancer screening: The gender dilemma. United European gastroenterology journal. 2017;5(3):448-54.

125.   Benedix F, Kube R, Meyer F, Schmidt U, Gastinger I, Lippert H, et al. Comparison of 17,641 patients with right- and left-sided colon cancer: differences in

epidemiology, perioperative course, histology, and survival. Diseases of the colon and rectum. 2010;53(1):57-64.

126.   Blom J, Lowbeer C, Elfstrom KM, Sventelius M, Ohman D, Saraste D, et al. Gender-specific cut-offs in colorectal cancer screening with FIT: Increased compliance and equal positivity rate. Journal of medical screening. 2019;26(2):92-7.

127.   Sarkeala T, Farkkila M, Anttila A, Hyoty M, Kairaluoma M, Rautio T, et al. Piloting gender-oriented colorectal cancer screening with a faecal immunochemical test: population-based registry study from Finland. BMJ open. 2021;11(2):e046667.

128.   Kapidzic A, van der Meulen MP, Hol L, van Roon AH, Looman CW, Lansdorp-Vogelaar I, et al. Gender Differences in Fecal Immunochemical Test Performance for Early Detection of Colorectal Neoplasia. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association. 2015;13(8):1464-71 e4.

129.   Alvarez-Urturi C, Andreu M, Hernandez C, Perez-Riquelme F, Carballo F, Ono A, et al. Impact of age- and gender-specific cut-off values for the fecal immunochemical test for hemoglobin in colorectal cancer screening. Digestive and liver disease: official journal of the Italian Society of Gastroenterology and the Italian Association for the Study of the Liver. 2016;48(5):542-51.

130.   Brenner H, Haug U, Hundt S. Sex differences in performance of fecal occult blood testing. The American journal of gastroenterology. 2010;105(11):2457-64.

131.   Gies A, Niedermaier T, Alwers E, Hielscher T, Weigl K, Heisser T, et al. Consistent Major Differences in Sex- and Age-Specific Diagnostic Performance among Nine Faecal Immunochemical Tests Used for Colorectal Cancer Screening. Cancers (Basel). 2021;13(14).

132.   Clark GR, Digby J, Fraser CG, Strachan JA, Steele RJ. Faecal haemoglobin concentrations in women and men diagnosed with colorectal cancer in a national screening programme. Journal of medical screening. 2021:9691413211056970.

133.   de Klerk CM, Vendrig LM, Bossuyt PM, Dekker E. Participant-Related Risk Factors for False-Positive and False-Negative Fecal Immunochemical Tests in Colorectal Cancer Screening: Systematic Review and Meta-Analysis. The American journal of gastroenterology. 2018;113(12):1778-87.

134.   Amitay EL, Cuk K, Niedermaier T, Weigl K, Brenner H. Factors associated with false-positive fecal immunochemical tests in a large German colorectal cancer screening study. International journal of cancer Journal international du cancer. 2019;144(10):2419-27.

135.   Stegeman I, de Wijkerslooth TR, Stoop EM, van Leerdam M, van Ballegooijen M, Kraaijenhagen RA, et al. Risk factors for false positive and for false negative test results

in screening with fecal occult blood testing. International journal of cancer Journal international du cancer. 2013;133(10):2408-14.

136.   Brenner H, Calderazzo S, Seufferlein T, Ludwig L, Dikopoulos N, Mangold J, et al. Effect of a Single Aspirin Dose Prior to Fecal Immunochemical Testing on Test Sensitivity for Detecting Advanced Colorectal Neoplasms: A Randomized Clinical Trial. Jama. 2019;321(17):1686-92.

137.   Lansdorp-Vogelaar I, Meester R, de Jonge L, Buron A, Haug U, Senore C. Risk-stratified strategies in population screening for colorectal cancer. International journal of cancer Journal international du cancer. 2022;150(3):397-405.

138.   Auge JM, Pellise M, Escudero JM, Hernandez C, Andreu M, Grau J, et al. Risk stratification for advanced colorectal neoplasia according to fecal hemoglobin concentration in a colorectal cancer screening program. Gastroenterology. 2014;147(3):628-36 e1.

139.   Stegeman I, de Wijkerslooth TR, Stoop EM, van Leerdam ME, Dekker E, van Ballegooijen M, et al. Combining risk factors with faecal immunochemical test outcome for selecting CRC screenees for colonoscopy. Gut. 2014;63(3):466-71.

140.   Kortlever TL, van der Vlugt M, Dekker E, Bossuyt PMM. Individualized faecal immunochemical test cut-off based on age and sex in colorectal cancer screening. Prev Med Rep. 2021;23:101447.

141.   Omata F, Shintani A, Isozaki M, Masuda K, Fujita Y, Fukui T. Diagnostic performance of quantitative fecal immunochemical test and multivariate prediction model for colorectal neoplasms in asymptomatic individuals. European journal of gastroenterology & hepatology. 2011;23(11):1036-41.

142.   Park CH, Jung YS, Kim NH, Park JH, Park DI, Sohn CI. Usefulness of risk stratification models for colorectal cancer based on fecal hemoglobin concentration and clinical risk factors. Gastrointestinal endoscopy. 2019;89(6):1204-11 e1.

143.   Cooper JA, Parsons N, Stinton C, Mathews C, Smith S, Halloran SP, et al. Risk-adjusted colorectal cancer screening using the FIT and routine screening data: development of a risk prediction model. British journal of cancer. 2018;118(2):285-93.

144.   Mikkelsen EM, Thomsen MK, Tybjerg J, Friis-Hansen L, Andersen B, Jorgensen JCR, et al. Colonoscopy-related complications in a nationwide immunochemical fecal occult blood test-based colorectal cancer screening program. Clinical epidemiology. 2018;10:1649-55.

145.   Kooyker AI, Toes-Zoutendijk E, Opstal-van Winden AWJ, Buskermolen M, van Vuuren HJ, Kuipers EJ, et al. Colonoscopy-Related Mortality in a Fecal Immunochemical Test-Based Colorectal Cancer Screening Program. Clinical

gastroenterology and hepatology: the official clinical practice journal of the American Gastroenterological Association. 2021;19(7):1418-25.

146.   Benazzato L, Zorzi M, Antonelli G, Guzzinati S, Hassan C, Fantin A, et al. Colonoscopy-related adverse events and mortality in an Italian organized colorectal cancer screening program. Endoscopy. 2021;53(5):501-8.

147.   Wieten E, Schreuders EH, Grobbee EJ, Nieboer D, Bramer WM, Lansdorp-Vogelaar I, et al. Incidence of faecal occult blood test interval cancers in population-based colorectal cancer screening: a systematic review and meta-analysis. Gut. 2019;68(5):873-81.

148.   Zorzi M, Hassan C, Senore C, Capodaglio G, Turrin A, Narne E, et al. Interval colorectal cancers after negative faecal immunochemical test in a 13-year screening programme. Journal of medical screening. 2021;28(2):131-9.

149.   Mancini S, Bucchi L, Giuliani O, Ravaioli A, Vattiato R, Baldacchini F, et al. Proportional incidence of interval colorectal cancer in a large population-based faecal immunochemical test screening programme. Digestive and liver disease : official journal of the Italian Society of Gastroenterology and the Italian Association for the Study of the Liver. 2020;52(4):452-6.

150.   Toes-Zoutendijk E, Kooyker AI, Dekker E, Spaander MCW, Opstal-van Winden AWJ, Ramakers C, et al. Incidence of Interval Colorectal Cancer After Negative Results From First-Round Fecal Immunochemical Screening Tests, by Cutoff Value and Participant Sex and Age. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association. 2020;18(7):1493-500.

151.   Lee CK, Choi KS, Eun CS, Park DI, Han DS, Yoon M, et al. Risk and Characteristics of Postcolonoscopy Interval Colorectal Cancer after a Positive Fecal Test: A Nationwide Population-Based Study in Korea. Cancer research and treatment : official journal of Korean Cancer Association. 2018;50(1):50-9.

152.   Chiu SY, Chuang SL, Chen SL, Yen AM, Fann JC, Chang DC, et al. Faecal haemoglobin concentration influences risk prediction of interval cancers resulting from inadequate colonoscopy quality: analysis of the Taiwanese Nationwide Colorectal Cancer Screening Program. Gut. 2017;66(2):293-300.

153.   Winawer SJ, Zauber AG, Ho MN, O'Brien MJ, Gottlieb LS, Sternberg SS, et al. Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. The New England journal of medicine. 1993;329(27):1977-81.

154.   Zauber AG, Winawer SJ, O'Brien MJ, Lansdorp-Vogelaar I, van Ballegooijen M, Hankey BF, et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. The New England journal of medicine. 2012;366(8):687-96.

155.   Thiis-Evensen E, Kalager M, Bretthauer M, Hoff G. Long-term effectiveness of endoscopic screening on incidence and mortality of colorectal cancer: A randomized trial. United European gastroenterology journal. 2013;1(3):162-8.

156.   Breekveldt ECH, Lansdorp-Vogelaar I, Toes-Zoutendijk E, Spaander MCW, van Vuuren AJ, van Kemenade FJ, et al. Colorectal cancer incidence, mortality, tumour characteristics, and treatment before and after introduction of the faecal immunochemical testing-based screening programme in the Netherlands: a population-based study. The lancet Gastroenterology & hepatology. 2022;7(1):60-8.

157.   Levin TR, Corley DA, Jensen CD, Schottinger JE, Quinn VP, Zauber AG, et al. Effects of Organized Colorectal Cancer Screening on Cancer Incidence and Mortality in a Large Community-Based Population. Gastroenterology. 2018;155(5):1383-91 e5.

158.   Larsen MB, Njor S, Ingeholm P, Andersen B. Effectiveness of Colorectal Cancer Screening in Detecting Earlier-Stage Disease-A Nationwide Cohort Study in Denmark. Gastroenterology. 2018;155(1):99-106.

159.   Ventura L, Mantellini P, Grazzini G, Castiglione G, Buzzoni C, Rubeca T, et al. The impact of immunochemical faecal occult blood testing on colorectal cancer incidence. Digestive and liver disease : official journal of the Italian Society of Gastroenterology and the Italian Association for the Study of the Liver. 2014;46(1):82-6.

160.   Chiu HM, Chen SL, Yen AM, Chiu SY, Fann JC, Lee YC, et al. Effectiveness of fecal immunochemical testing in reducing colorectal cancer mortality from the One Million Taiwanese Screening Program. Cancer. 2015;121(18):3221-9.

161.   Lee YC, Li-Sheng Chen S, Ming-Fang Yen A, Yueh-Hsia Chiu S, Ching-Yuan Fann J, Chuang SL, et al. Association Between Colorectal Cancer Mortality and Gradient Fecal Hemoglobin Concentration in Colonoscopy Noncompliers. Journal of the National Cancer Institute. 2017;109(5).

162.   Zorzi M, Fedeli U, Schievano E, Bovo E, Guzzinati S, Baracco S, et al. Impact on colorectal cancer mortality of screening programmes based on the faecal immunochemical test. Gut. 2015;64(5):784-90.

163.   Idigoras Rubio I, Arana-Arri E, Portillo Villares I, Bilbao Iturribarrria I, Martinez-Indart L, Imaz-Ayo N, et al. Participation in a population-based screening for colorectal cancer using the faecal immunochemical test decreases mortality in 5 years. European journal of gastroenterology & hepatology. 2019;31(2):197-204.

164.   Keys MT, Serra-Burriel M, Martinez-Lizaga N, Pellise M, Balaguer F, Sanchez A, et al. Population-based organized screening by faecal immunochemical testing and colorectal cancer mortality: a natural experiment. International journal of epidemiology. 2021;50(1):143-55.

165.   Zhong GC, Sun WP, Wan L, Hu JJ, Hao FB. Efficacy and cost-effectiveness of fecal immunochemical test versus colonoscopy in colorectal cancer screening: a systematic review and meta-analysis. Gastrointestinal endoscopy. 2020;91(3):684-97 e15.

166.   Aronsson M, Carlsson P, Levin LA, Hager J, Hultcrantz R. Cost-effectiveness of high-sensitivity faecal immunochemical test and colonoscopy screening for colorectal cancer. The British journal of surgery. 2017;104(8):1078-86.

167.   Areia M, Fuccio L, Hassan C, Dekker E, Dias-Pereira A, Dinis-Ribeiro M. Cost-utility analysis of colonoscopy or faecal immunochemical test for population-based organised colorectal cancer screening. United European gastroenterology journal. 2019;7(1):105-13.

168.   Meulen MPV, Kapidzic A, Leerdam MEV, van der Steen A, Kuipers EJ, Spaander MCW, et al. Do Men and Women Need to Be Screened Differently with Fecal Immunochemical Testing? A Cost-Effectiveness Analysis. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2017;26(8):1328-36.

169.   Thomas C, Mandrik O, Whyte S, Saunders CL, Griffin SJ, Usher-Smith JA. Should colorectal cancer screening start at different ages for men and women? Cost-effectiveness analysis for a resource-constrained service. Cancer Rep (Hoboken). 2021;4(4):e1344.

170.   Registered clinical trial identifier NCT02078804 available from: https://clinicaltrials.gov/ct2/show/NCT02078804 cited in Nov 2021

171.   Moberger P, Skoldberg F, Birgisson H. Evaluation of the Swedish Colorectal Cancer Registry: an overview of completeness, timeliness, comparability and validity. Acta oncologica. 2018;57(12):1611-21.

172.   The Board of Health and Welfare, Nord-DRG database on reference costs in hospitals and clinics for diagnoses and procedures, available from https://www.socialstyrelsen.se/utveckla-verksamhet/e-halsa/klassificering-och-koder/drg/viktlistor/ cited on Nov 2021

173.   The National Board of Health and Welfare, cancer register available from https://www.socialstyrelsen.se/statistik-och-data/register/alla-register/cancerregistret/ cited on Nov 2021

174.   Ghisletta P, Spini D. An Introduction to Generalized Estimating Equations and an Application to Assess Selectivity Effects in a Longitudinal Study on Very Old Individuals. Journal of Educational and Behavioral Statistics 2004;29(4):421-437 doi: 10.3102/10769986029004421

175.	Anderson BW, Smyrk TC, Anderson KS, Mahoney DW, Devens ME, Sweetser SR, et al. Endoscopic overestimation of colorectal polyp size. Gastrointestinal endoscopy. 2016;83(1):201-8.

176.	van Putten PG, Hol L, van Dekken H, Han van Krieken J, van Ballegooijen M, Kuipers EJ, et al. Inter-observer variation in the histological diagnosis of polyps in colorectal cancer screening. Histopathology. 2011;58(6):974-81.

177.	Koskenvuo L, Malila N, Pitkaniemi J, Miettinen J, Heikkinen S, Sallinen V. Sex differences in faecal occult blood test screening for colorectal cancer. The British journal of surgery. 2019;106(4):436-47.

178.	Pitkaniemi J, Seppa K, Hakama M, Malminiemi O, Palva T, Vuoristo MS, et al. Effectiveness of screening for colorectal cancer with a faecal occult-blood test, in Finland. BMJ Open Gastroenterol. 2015;2(1):e000034.

179.	Shaukat A, Mongin SJ, Geisser MS, Lederle FA, Bond JH, Mandel JS, et al. Long-term mortality after screening for colorectal cancer. The New England journal of medicine. 2013;369(12):1106-14.

180.	Libby G, Fraser CG, Carey FA, Brewster DH, Steele RJC. Occult blood in faeces is associated with all-cause and non-colorectal cancer mortality. Gut. 2018;67(12):2116-23.

181.	Cox B, Sneyd MJ. Bias in breast cancer research in the screening era. Breast. 2013;22(6):1041-5.

182.	Marcus PM, Prorok PC, Miller AB, DeVoto EJ, Kramer BS. Conceptualizing overdiagnosis in cancer screening. Journal of the National Cancer Institute. 2015;107(4).

183.	Chiu HM, Jen GH, Wang YW, Fann JC, Hsu CY, Jeng YC, et al. Long-term effectiveness of faecal immunochemical test screening for proximal and distal colorectal cancers. Gut. 2021;70(12):2321-9.

184.	Jover R, Zapater P, Bujanda L, Hernandez V, Cubiella J, Pellise M, et al. Endoscopist characteristics that influence the quality of colonoscopy. Endoscopy. 2016;48(3):241-7.

185.	Schachschal G, Sehner S, Choschzick M, Aust D, Brandl L, Vieth M, et al. Impact of reassessment of colonic hyperplastic polyps by expert GI pathologists. International journal of colorectal disease. 2016;31(3):675-83.

186.	Heinavaara S, Gini A, Sarkeala T, Anttila A, de Koning H, Lansdorp-Vogelaar I. Optimizing screening with faecal immunochemical test for both sexes - Cost-effectiveness analysis from Finland. Preventive medicine. 2022:106990.