from FIT to Future

ADVANCES IN COLORECTAL CANCER SCREENING

Esmée Grobbee

from FIT to Future

advances in colorectal cancer screening

Esmée Grobbee

Colofon

The work presented in this thesis was conducted at the department of Gastroenterology and Hepatology, Erasmus MC University Medical Center, Rotterdam, The Netherlands.

Financial support for printing this thesis was kindly provided by the Department of Gastroenterology and Hepatology, Erasmus MC University Medical Center, Rotterdam; Nederlandse Vereniging voor Gastroenterologie; Ferring B.V.; Bevolkingsonderzoek Zuid-West; Zambon Nederland B.V. ; Olympus Nederland B.V.; Dr. Falk Pharma Benelux B.V.; Chipsoft B.V.; Tramdedico B.V.; DaklaPack Clinical Trials B.V.

from FIT to Future

advances in colorectal cancer screening

FIT naar de toekomst

ontwikkelingen op het gebied van darmkankerscreening

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

prof.dr. H.A.P. Pols

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
vrijdag 4 november 2016 om 13.30 uur

Elisabeth Jacomina Grobbee

Esmée

geboren te Nieuwegein

Erasmus University Rotterdam

Promotiecommissie

Promotor:

Prof. dr. E.J. Kuipers

Overige leden:

Prof. dr. M.J. Bruno

Prof. dr. P.M.M. Bossuyt

Prof. dr. J.F. Lange

Copromotor:

Dr. M.C.W. Spaander

Contents

Wijsheid begint met verwondering

Plato

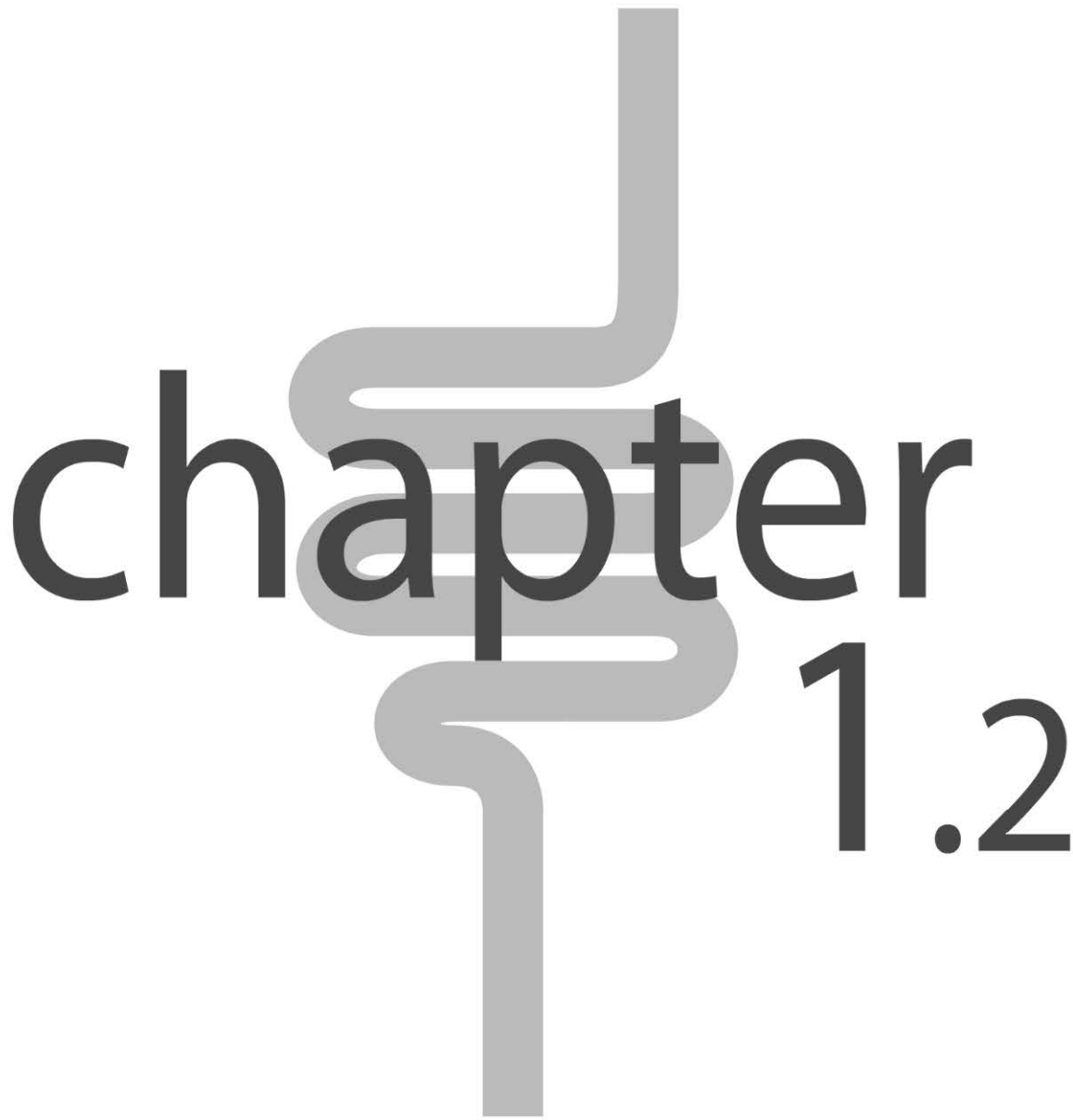
PART I

general introduction

chapter 1.1

General introduction

Colorectal cancer (CRC) is the second most common malignancy, as well as the second most common cause of cancer-related death in the Western world[1]. CRC has a long pre-clinical phase involving adenomas that slowly progress over time into carcinoma. This long pre-clinical phase, combined with the fact that CRC is often symptomless in the early stages, make the disease very suitable for screening[2,3]. Many screening methods are currently available and can be broadly divided into invasive imaging strategies and non-invase strategies. The latter are in particular based on fecal occult blood testing (FOBT), but other fecal as well blood tests are available. Current European guidelines recommend fecal immunochemical testing (FIT)[4]. This thesis will explore the current status of colorectal cancer screening with focus on FIT. Part one provides a discussion of the main topics of this thesis. The second part will explore different screening strategies and the effect of FIT over multiple rounds of screening. In the third part the use of FIT as a quantitative test will be further evaluated. In the fourth part quality issues concerning CRC screening will be investigated. In the last part, findings of this thesis will be discussed and future perspectives will be provided.


chapter
1.2

Short introduction on colorectal cancer screening

infographic

Colorectal

Colorectal cancer (CRC) is the #2 second most common malignancy world-wide, as well as the second most common cause of cancer-related death. Its incidence is rising rapidly, both as a result of expansion of the elderly population as well as an increase in risk factors, such as a Western diet, a sedentary lifestyle, smoking and alcohol-intake

Colorectal cancer meets all the criteria as defined by Wilson and Jugner in 1968:

CRC

- has a high MORBIDITY and MORTALITY
- is PREVENTABLE and EARLY detectable
- screening TESTS are available
- is TREATABLE with good outcomes when detected in an early stage

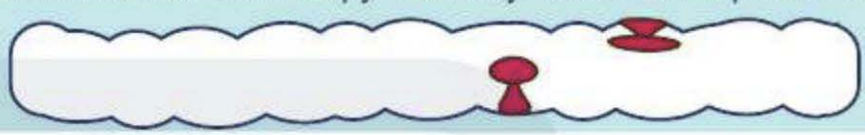
SCREENING

CRC has a higher incidence in men than in women

FIT is less sensitive for detection of CRC in women

50+ After the age of 50 CRC incidence increases

Polyps, i.e. adenomas, are precursors of CRC, in the so-called adenoma-carcinoma sequence.These adenomas can be detected and removed at colonoscopy. This way CRC can be prevented.



Current strategies can broadly be divided in to two: colonic imaging techniques or fecal occult blood tests

CURRENT STRATEGIES

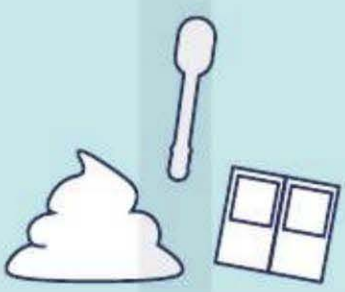
COLONIC IMAGING TECHNIQUES

Endoscopic screening using colonosocpy is considered the gold standard. Other screening strategies include sigmoidoscopy, and colonic imaging by computed-tomography colonography

FECAL OCCULT BLOOD TESTS

Fecal occult blood based tests (FOBT) are non-invase and most commonly used are the guaiac FOBT and fecal immunochemical test (FIT). At present, FIT is the preferred method of FOBT-based sceening

FOBT screening reduces CRC-related mortality but not overall mortality



Cancer

FUTURE STRATEGIES

An increasing number of countries are commencing on, or are already screening for, colorectal cancer. As CRC screening progresses, new questions arise regarding optmizing current screening strategies and the development of novel methods.

At present FIT is the preferred method of screening and recommended by European guidelines. However, many questions still remain, such as:

Is FIT really better than guaiac FOBT to screen for colorectal cancer?

Is FIT more effective than endoscopic strategies over multple rounds?

What is the effect of FIT screening over multiple rounds on attendance and diagnostic yield?

Can fecal hemoglobin (Hb) concentration be used in risk scores or to alter screening intervals?

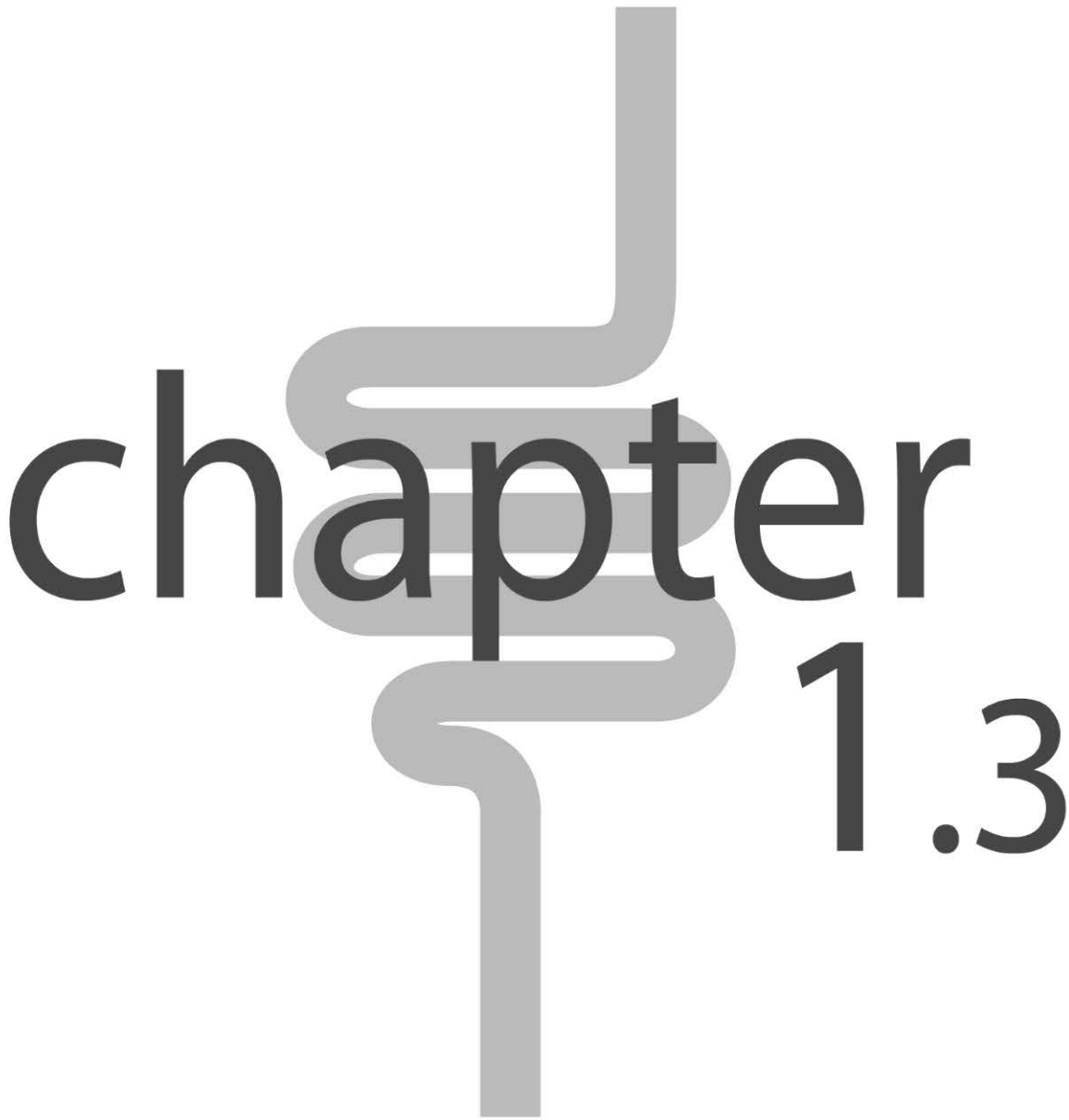
Can FIT be combined with biomarkers such as DNA and the gut microbiome?

What is the quality of colonoscopy and are early CRCs (i.e., T1 lesions) adequately recognized and treated?

QA? This thesis will try to answer these and other questions by exploring the current status of colorectal cancer screening and novel developements

Quality is an important issue in CRC screening and mainly involves quality of colonoscopy. Multiple quality indicators have been established, most notable are the adenoma detection rate (ADR) and cecal intubation rate (CIR). An increased ADR is associated with less interval CRCs. Lasty, cleanliness of the bowel is an important issue and can be measuered by the Boston Bowel Preparation Score.

QUALITY


chapter
1.3

Current status of colorectal cancer screening

adapted from:

Esmée J. Grobbee, Eline H. Schreuders, Ernst J. Kuipers. Are there effective screening programs in Europe?
Multidisciplinary management of rectal cancer 2nd edition, 2016 - in press

Esmée J. Grobbee, Eline H. Schreuders, Manon C.W. Spaander, Ernst J. Kuipers. Colorectal cancer screening and diagnosis.
Textbook of hepato-gastroenterology. Part II: gastroenterology, 2016 - in press

Eline H. Schreuders, Esmée J. Grobbee, Manon C.W. Spaander, Ernst J. Kuipers, Advances in fecal tests for colorectal cancer screening
Current Treatment Options Gastroenterology, 2016

Introduction

Colorectal cancer (CRC) is the second most common malignancy, as well as the second most common cause of cancer-related death in the European Union[5]. The number of new cases is approximately 430,000 per year. This number is rising rapidly, both as a result of expansion of the elderly population as well as an increase in risk factors, including a sedentary life style, smoking, alcohol intake and obesity. Around 45% of patients diagnosed with CRC will die as a result of the disease in spite of intensive treatment[6].

Patients are diagnosed with CRC either through screening or when they present with symptoms. Colorectal cancer develops from adenomas that are seen as non-malignant precursors of CRC. A small proportion of adenomas slowly progress over time into carcinoma; the so-called adenoma-carcinoma sequence[3,7]. This pre-clinical stage is often symptomless, which explains that CRC mostly remains undetected at an early stage. Symptoms related to CRC include a change in bowel habits, abdominal pain, weight loss, blood loss, anemia and fatigue. Less frequent symptoms as first presentation of CRC are symptoms of obstruction, or metastatic disease such as peritonitis or jaundice. At present, colonoscopy is considered the gold standard for CRC diagnosis in symptomatic patients. Colonoscopy allows visualization of the entire colon, as well as tissue sampling for histopathology, and instantaneous removal of (precancerous) polyps (Figure 1).

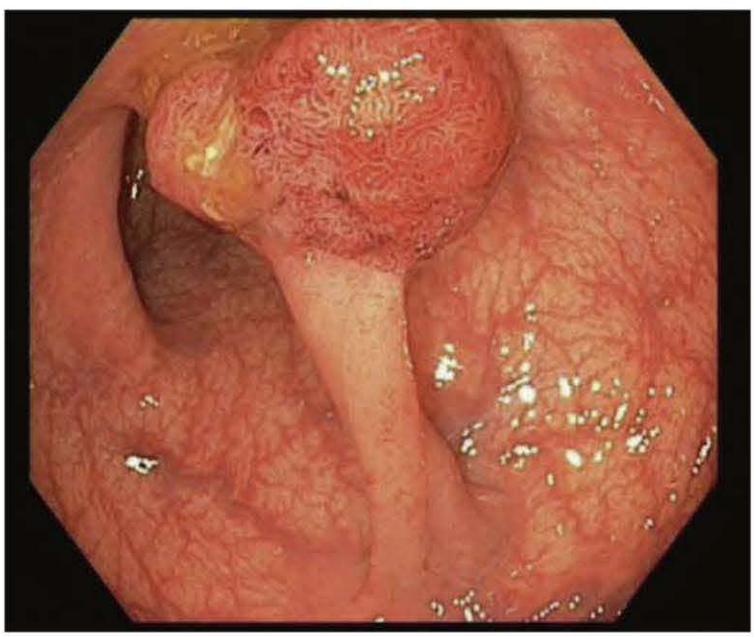


Figure 1. *Pedunculated advanced adenoma at colonoscopy*

Screening for colorectal cancer

Screening is an appealing concept for colorectal cancer as the disease meets all the screening criteria as defined by Wilson and Jungner in 1968[8]. These screening principles, later redefined by the WHO, state that screening is justified when a disease is common and associated with a high morbidity and mortality, with availability of adequate screening tests to detect the disease at an early stage with the possibility of treatment with improved outcome compared to later detection[9]. Colorectal cancer is indeed common, has a long pre-clinical phase, and often does not cause symptoms until at a late stage. In addition, when CRC is detected at an early stage, various treatment options are available and survival is considerably higher than at an advanced stage of disease[10]. All these factors combined make screening for CRC more suitable than any other malignancy[2].

Several screening methods are available varying widely in invasiveness and diagnostic accuracy. Screening modalities can broadly be divided into invasive imaging screening strategies and non-invasive screening strategies. The latter are in particular based on fecal testing, but novel strategies are being developed based on breath and blood tests. No single screening test has been shown to be advantageous over others with respect to impact on CRC-related morbidity and mortality. Many screening strategies involve a two-step method, starting with one screening test followed by colonoscopy in those who test positive. This approach is generally associated with higher participation than primary colonoscopy screening, and has the advantage of requiring fewer endoscopic resources[11]. The choice of a screening test is most often based on the available colonoscopy resources, expected participation rates, and intended detection rates. It is important to realize that all screening strategies, except colonoscopy, require a follow-up colonoscopy in case of a positive exam or test. Furthermore, each strategy involves a different screening interval at which the procedure needs to be repeated in case of a negative previous result.

Endoscopic screening strategies

Colonoscopy is considered the gold standard for CRC diagnosis. Therefore, many consider primary colonoscopy screening as the preferred method for screening. The fact that colonoscopy allows complete visualization of the colon and simultaneous removal of polyps is an important advantage of colonoscopy-based screening. This enables reduction of both incidence and mortality of CRC[3]. Nevertheless, there are also disadvantages to colonoscopy screening such as the significant burden and costs[12]. Consequently, participation rates in colonoscopy-based screening programs tend to be

low, often resulting in a low number of detected neoplasia relative to the invited population(13). Moreover, because of the invasiveness of the procedure, there is a risk of serious complications such as haemorrhage and perforation[12]. A colonoscopy is considered to be required every 10 years in case of no findings at screening colonoscopy (i.e. a negative colonoscopy).
Flexible sigmoidoscopy (FS) examines the colon up to the splenic flexure. Screenees are usually referred for subsequent colonoscopy when polyps larger than 10 mm or more than three polyps are seen. Concerns have been raised about missing right-sided lesions, and the resulting limited efficacy in reducing proximal CRC mortality[12]. However, a decrease in CRC incidence and mortality has been shown up to 23% and 31%, respectively[14]. Most often used screening interval for FS is 5 years in case no (advanced) polyps are found at endoscopy.

Computed tomography-colonography (CTC) allows radiological imaging of the entire colon in two-dimensional and three-dimensional images. In case large polyps are detected at CTC (usually defined as polyps with a diameter ≥ 6mm or ≥ 10mm) patients are referred for colonoscopy. CTC tends to generate higher participation rates than colonoscopy[15]. In a randomized trial comparing CTC with colonoscopy a similar yield for advanced neoplasia amongst invitees was found, yet this finding was mainly due to the relatively high participation rate of CTC[13]. Similar to FS, a CTC is to be repeated every 5 years in case no significant findings are found.

Video capsule endoscopy (VCE) is the most recently emerging endoscopic screening strategy[16]. It involves the use of a wireless camera shaped in the form of a large capsule that is swallowed by the screenee. It allows visualization of the entire gastrointestinal tract. In the past years VCE has undergone considerable improvement regarding frame speed and angle of view resulting in a sensitivity up to 88% and a specificity of 82% for the detection of adenomas >5 mm in population-based screening setting[17].

Fecal occult blood test screening strategies

While adenomas don't cause overt symptoms in most subjects, they can intermittently bleed which can lead to detection of occult blood in feces. Essentially, two types of fecal occult blood tests (FOBTs) are available: the guaiac FOBT and fecal immunochemical test (FIT). The gFOBT detects blood by using paper from Guaiacum trees and hydroperoxidase. This leads to a blue discoloration of the paper when haem is present. A positive result of the test is defined by this blue discoloration. A test usually consists of three sets of cards with two panels each. The test needs to be performed

three times by the screenee on separate stools. Several, large, randomized trials have demonstrated that screening with repeated gFOBT reduces CRC-related mortality[18,19]. At present the use of gFOBT is largely replaced by FIT for several reasons. Firstly, FIT is easier to use than gFOBT and requires the screenee to take only one stool sample resulting in higher participation rates[20]. Secondly, FIT has a higher sensitivity compared to gFOBT in particular for the detection of precancerous adenomatous polyps. FIT is consequently associated with a higher detection rate of advanced neoplasia per screening round. This results in a lower rate of interval carcinomas[21]. Lastly, FITs detect human globin using an antibody-based assay. This allows automated, quantitative measurements, whereas gFOBT is only available as a qualitative test that needs to be read manually. For all the above-mentioned reasons, FIT is currently regarded as the preferred method of non-invasive screening and recommended by European as well AsiaPacific guidelines[4,22]. Quantitative FIT output allows for adjusting the threshold for the definition of a positive test, which is relevant in situations with limited colonoscopy capacity as the cut-off can be adjusted to yield maximal results within restricted resources. Many FIT brands are available worldwide, that all use varying sampling techniques (e.g. brushes or spatulas) and buffers (Figure 2).

Figure 2. *Different FIT brands.*

It is becoming more evident that different brands of FIT vary in measurement results leading to different positivity rates at the same cut-off expressed in µg Hb/g feces (23, 24). So far, there is no evidence favouring one FIT over another[12,25]. Screenees undergoing gFOBT or FIT are recommended to repeat screening annually or biennially.

Novel developments in non-invasive screening strategies

In the past years many novel screening modalities have emerged rapidly. These are all relatively new compared to established endoscopic and FOBT-based screening strategies. These screening modalities are most often non-invasive strategies involving DNA/RNA or protein biomarkers, the fecal microbiome and volatile markers. Biomarkers are based on the principle that malignant lesions shed cells or blood into the feces that can be detected in the form of aberrant DNA, RNA or proteins[26]. It is of importance to realize that the performance of such biomarkers might be different per CRC subtype as CRC can develop through various pathways. This implicates the need for the use of multiple tumor markers[27,28]. A recent study using a multi-target stool DNA and FIT showed a marked improvement in sensitivity when combined with FIT, mainly regarding the detection of advanced adenomas (42% using the combination versus 23% for FIT alone)[29]. It is notable that this increase in sensitivity was joined by a much higher positivity rate requiring more colonoscopy resources and a decrease in specificity. Regarding protein-based biomarkers, fecal tumor M2 pyruvate kinase (M2-PK) has received much attention, with a reported sensitivity and specificity of 79% and 80% for CRC[30]. However, to detect these biomarkers, often a full stool sample is needed, which comes with much impracticality and requires a considerable effort from the screenees. While such inconveniences could substantially influence adherence to a screening program, this has not yet been investigated and requires further research.

Growing attention has been given to gut microbiota in colorectal cancer development, under the assumption that the presence of colorectal neoplasia may be associated with specific microbiota[31-34]. One of the first bacteria associated with CRC was Streptococcus bovis[35]. At present, several bacteria have been recognized to be of importance in colorectal carcinogenesis, such as E. Coli, Bacteroides Fragilis and Fusobacterium nucleatum[33,36]. The exact role of these bacteria has yet to be elucidated, however, it is thought that multiple bacteria play different roles and that carcinogenesis is a dynamic process associated with diverse changes in the microbiome[37]. Combining microbiota with current non-invasive screening strategies could theoretically improve screening strategies as it could also detect non-bleeding lesions that are missed using FOBT-based strategies.

Volatile organic biomarkers could present an interesting new possibility for the detection and screening of CRC. Volatile organic compounds can be found in various excreted biological materials, such as feces, urine, and breath[38]. It has been shown that these volatile markers can be used as a 'smell-print' for CRC and are detectable by the use of canine scent detection or electronic nose (39). However, analyzing volatile markers is at present costly and inconvenient, as it requires the

use of dogs or nanosensors. Following, research on the use of these markers both in diagnosis of CRC in symptomatic patients as well as in a screening setting is sparse and further research is much awaited.

Quality of colonoscopy in colorectal cancer screening

Given the low point-prevalence of CRC, measuring cancer detection rates is not feasible as a quality indicator for CRC screening programs, resulting in a need for surrogate indicators. Therefore, quality in colonoscopy is most often measured by adenoma detection rates (ADR) and cecal intubation rates (CIR), because they are considered to be a proxy for a thorough and complete examination of the colon[40,41] . The ADR is defined as the proportion of (screening) colonoscopies where at least one adenoma is found. Much focus has been on ADR, as an endoscopists' ADR is inversely correlated with post-colonoscopy interval cancer rates[42-44]. This supports the need to monitor endoscopists' ADRs. Most guidelines recommend an ADR equal to or above 25 % (30% in men and 20% in women)[44,45]. In the past years, many improvements have been made with respect to visualization of the colorectal mucosa, such as optimizing bowel preparation and development of higher resolution and wider view endoscopes[46]. Cleanliness of the colon is frequently scored by means of the validated Boston Bowel Preparation Score, to measure bowel preparation as a quality outcome among hospitals or within studies[47]. Studies have shown a correlation between adequate bowel preparation and ADR[46,48]. There is much evidence favoring a "split"-dose preparation, with the second dose taken on the same day as the colonoscopy. Increasing the amount of time spent viewing during the withdrawal of the scope has also been shown to affect the number of polyps found and increase the potential to detect adenomas. An endoscope withdrawal time of six to nine minutes is associated with a higher ADR compared to a withdrawal time of less than six minutes[49,50]. Not surprisingly, ADR is influenced by the quality of endoscopic imaging, which has greatly evolved over the past years. High-definition white light (HDWL) endoscopy has become the current standard and is recommended by the recent guideline[51]. There have been many studies focusing on novel image-enhancement techniques such as narrow-band imaging and auto-fluorescence endoscopy, and devices such as caps on the tip of the endoscope. No studies have convincingly shown superiority of use of any of these techniques over HDWL endoscopy in terms of ADR[2,46]. However, chromoendoscopy (i.e. adding through-the scope-infusions to improve mucosal visibility) has shown efficacy in the detection of small or flat mucosal lesions[52]. Notably, an improvement in detection of adenomas, may not necessarily result in a higher ADR and comes with substantial costs. Also, there are some limitations regarding practicality and more research is warranted to confirm if these utilities are of additional value in

adenoma detection and ultimately in decreasing CRC incidence.

In addition to technical improvements in colonoscopy equipment, improvement in optical diagnosis of polyps has received much attention. This refers to reliable endoscopic assessment of a polyp during colonoscopy, which would in theory make the pathologist's assessment unnecessary. Accurate optical diagnosis of small and diminutive polyps would allow to resect and discard these polyps, without the need for histopathological assessment[53]. Furthermore, with the worldwide introduction of screening programs, the proportion of CRCs that are diagnosed at an early stage increases[11]. The rise in the detection of these early stage lesions (i.e. T1 lesions) underlines the importance of adequate diagnosis and resection[54]. Several polyp classification systems have therefore been developed, to aid clinicians in recognizing worrisome features of polyps before resection[55-59]. Figure 3 shows a simplified overview of two frequently used classification systems, the Kudo pit pattern and NICE classification. The timely recognition and endoscopic removal of malignant polyps may reduce the need for surgery. Moreover, removal of precancerous adenomas and early stage cancers has a substantial effect on CRC mortality[3].

KUDO		FINDINGS	NICE	
Type I	round pits	normal mucosa		
Type II	asteroid or papillary pits	hyperplastic polyp	color: same or lighter than background vessels: none surface pattern: dark or white spots of same size or regular absence of pattern	Type 1
Type IIIs	small tubular or roundish pits	adenoma	color: browner than background vessels: brown vessels surface pattern: oval, tubular or branched white structures	Type 2
Type IIIL	large tubular or roundish pits			
Type IV	branch- or gyrus-like pits			
Type V	non-structural pits or loss of pits	submucosal carcinoma	color: brown to darker than background vessels: areas of disrupted or missing vessels surface pattern: unstructured or absent pattern	Type 3

Figure 3. *Simplified overview of KUDO and NICE classification.*

In case CRC is diagnosed, it is staged according to the Tumor Node Metastases (TNM) staging system of the American Joint Committee on Cancer/Union for International Cancer Control[60]. The clinical

staging (named cT, cN, cM) can be done through endoscopic, radiographic and intraoperative findings. Prior to surgical interventions the patient is most often referred for a CT-scan of the thorax and abdomen to assess local tumor size and invasions, lymphatic and distant metastases, and tumor-related complications. In addition, histological examination of the resection specimen is required to confirm the diagnosis of CRC, and for pathologic staging (named pT, pN, pM). TNM-staging of CRC provides an outline for therapy and prognosis. Furthermore, when the diagnosis of CRC is confirmed by histology, it is important to assess family history of cancer. It is estimated that in a relatively large number of patients a definable inherited component might play an important role in the development of CRC, such as for example Lynch syndrome[61]. Recent guidelines recommend to test for Lynch syndrome, including assessing microsatellite instability, by all CRC patients up to the age of 70 years old[62].

Current status of screening in Europe and the Netherlands

In 2003, the Council of the European Union recommended population screening for CRC. This led in 2010 to the publication of European Guidelines for Quality Assurance in Colorectal Cancer Screening and Diagnosis[4]. These guidelines outline targets for key performance indicators for CRC screening including adherence, follow-up, and cancer detection rates. In Europe large variations between screening programs exist for several reasons. Firstly, many programs were already ongoing at the time of publication, and therefore did not follow guidelines for standardization across Europe. Secondly, the guidelines leave some room for choice of screening strategy such as age ranges and screening intervals. Lastly, there are considerable differences with respect to financial resources and colonoscopy capacity across European countries.

In Europe, CRC screening primarily concerns organized population-based screening, which involves a systematic process of inviting the target population, including monitoring and often program quality assurance. This is in contrast with opportunistic screening, which happens when either an individual asks a treating physician for a screening test, or vice versa (screening is offered by a doctor or health professional). Unlike an organized screening program, opportunistic screening may not be monitored. In 1972 Germany was among one of the first countries to offer CRC screening by means of opportunistic gFOBT-based screening. In 2015, 24 European countries have established or are preparing a nationwide organized or opportunistic CRC screening program[63]. Of these 24 countries, 12 have a population-based organized program, nine have opportunistic programs, and three are executing pilot studies. Most countries with a high incidence have some form of screening in place. However, Slovakia has the highest CRC rates in Europe (age-standardized incidence rate is

42.7 per 100.000 and age-standardized mortality rate 18), without an organized screening program, but offers only opportunistic colonoscopy[63,64].

In the Netherlands, a screening program was started in January 2014, and is a biennial, FIT-based, organized population-based program. All inhabitants in the age 55 to 75 years are invited for screening with the FOB-Gold. There has been controversy about the type of FIT that was chosen, as previous research on CRC screening in the Netherlands was done using the OC-sensor[65].

Remarkable differences with regard to screening modality are seen between countries[63]. Some countries still use gFOBT, Italy uses FIT except for some regions where FS is used, and Germany uses gFOBT accompanied by opportunistic colonoscopy. Poland offers colonoscopy every ten years, and while initially having an opportunistic character, it changed into an organized program several years ago. Most other countries screen by means of FIT. The programs vary in their test cut-off as well as age range, but all screen men and women alike. They also, almost without exception, use a single test per screening round, and predominantly use two-year intervals. Large differences in adherence and colonoscopy uptake after a positive FIT exist. The choice for a particular cut-off is primarily determined by colonoscopy capacity. All these differences lead to different effects of screening in terms of cumulative yield, and possibly mortality. It is key to realize these differences in use of FIT exist. Also, it is important for countries to report their screening methods and outcomes to evaluate how these different FIT-programs relate to each other, and to be able to make a fair comparison.

Summary and conclusion

Colorectal cancer is a major burden worldwide, however incidence and mortality rates fluctuate markedly per country. Patients are either diagnosed when presenting with symptoms or through screening programs. Efficacy of a screening program depends on the willingness to undergo the screening modality as well as on test accuracy and burden. Additionally, as various screening methods are available, costs and colonoscopy capacity are of influence when implementing a screening program. In the past decennia many countries have commenced on a CRC screening program, though the screening strategies vary per country, with FOBT-based screening and colonoscopy being the most frequently used modalities. There is a shift from opportunistic screening towards organized population screening programs that contain extensive monitoring. In addition, previous gFOBT-based programs are now switching to FIT. Little evidence is yet available on the comparison of FIT-based programs and colonoscopy-based programs. Improvements could still be made in reporting and standardizing outcomes of screening programs, including the reporting of fecal hemoglobin concentrations for FIT. Moreover, despite the quantitative nature of the FIT, it is

still used invariably in a qualitative manner using a pre-specified cut-off. Future research on the effect of FIT-screening over multiple rounds, the occurrence of interval cancers, and ultimately its effect on incidence and mortality reduction is much awaited.

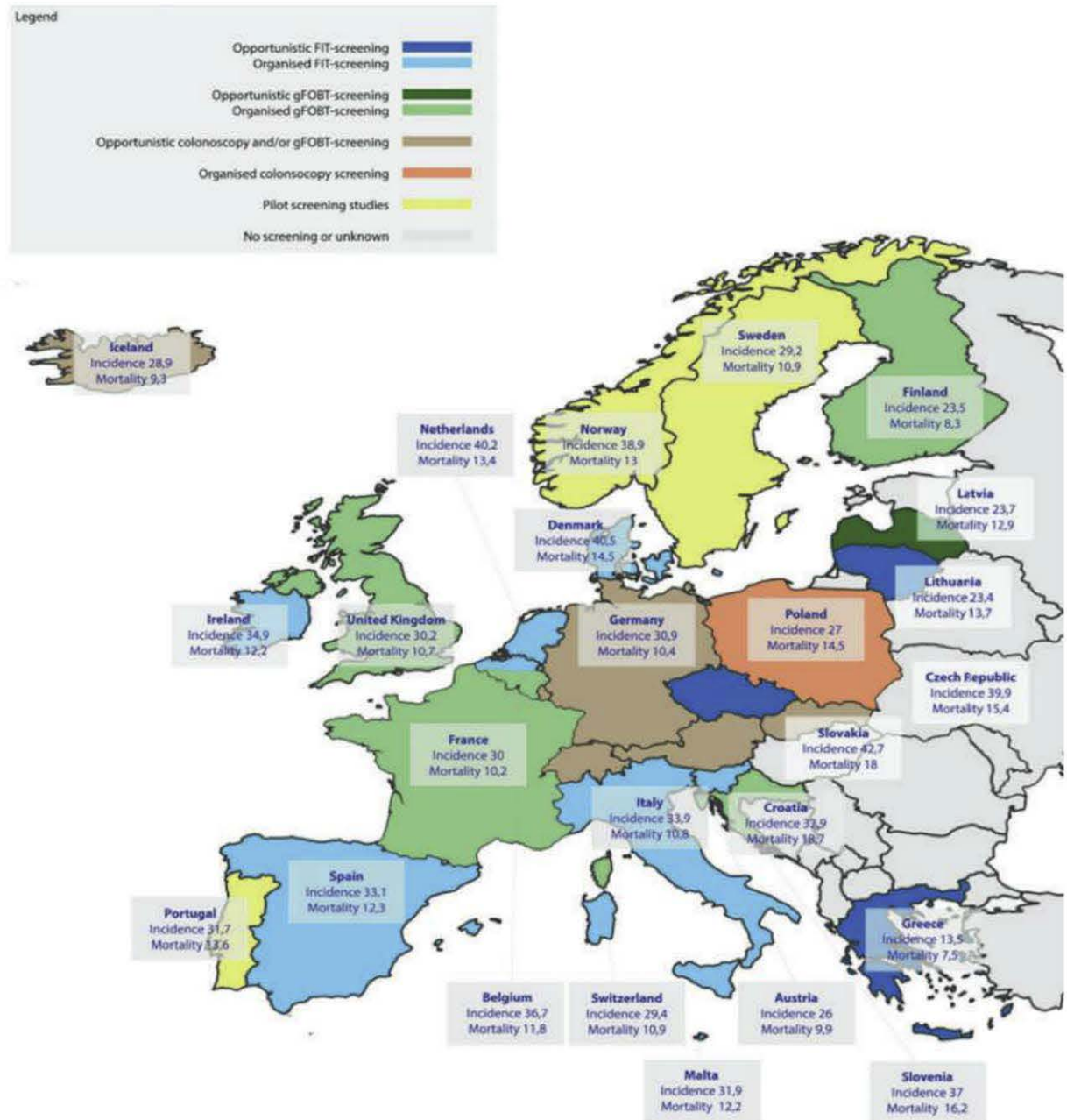
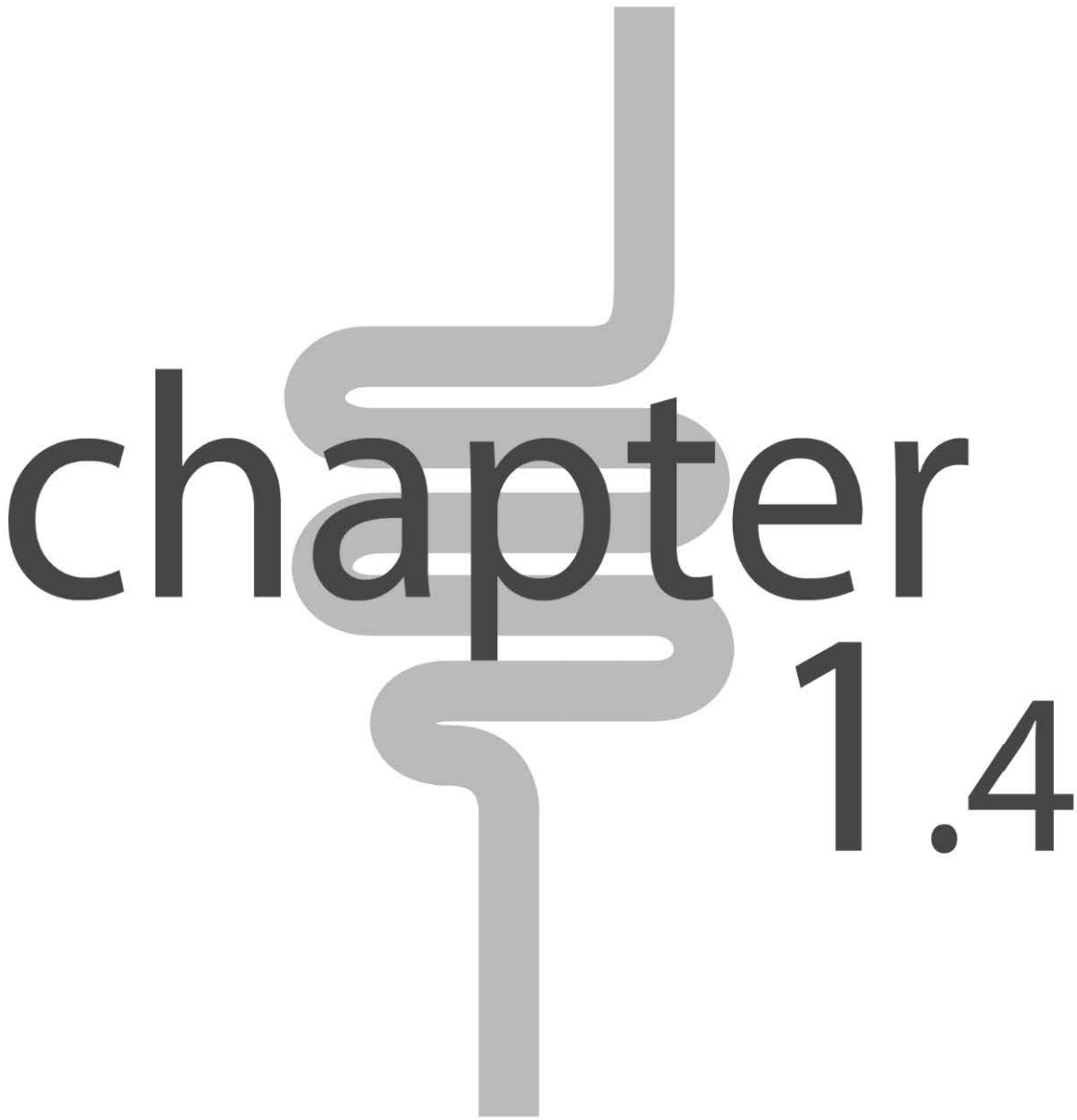


Figure 4. *Overview of age-standardized incidence and mortality rates, and screening programs in Europe (estimates by Globocan 64, original empty map retrieved from www.freevectormaps.com).*


chapter
1.4

Aims & outline of the thesis

Aims

In this thesis we aim to explore the current status of colorectal cancer screening, thereby comparing available screening methods, while focusing on the use of fecal immunochemical testing (FIT) in an organized population-based screening program.

Outline of the thesis

This thesis will start with describing the current status of colorectal cancer (CRC) screening both in the Netherlands as well as internationally (**PART I**).

In **PART II**, different screening modalities will be discussed with a focus on screening with FIT. At present many screening modalities are available and there is no evidence showing superiority of one screening strategy over another. At present, the most commonly used strategies can be broadly divided into non-invasive tests, such as fecal occult blood tests and invasive screening strategies, such as colonoscopy. In **Chapter 2** we will compare fecal occult blood testing with FIT versus guaiac fecal occult blood testing (gFOBT) in a Cochrane systematic review and meta-analysis. Following in **Chapter 3**, FIT will be compared to endoscopic screening strategies (sigmoidoscopy and colonoscopy) in a randomized setting over four rounds of FIT-screening. In **Chapter 4**, we will further examine FIT-based CRC screening by comparing two different FIT-brands and propose a new approach to adequately compare differences between brands. Next, as FIT screening requires biennial screening rounds, we will concentrate on multiple rounds of FIT screening and the effect on diagnostic yield after three rounds (**Chapter 5**). In **Chapter 6** we will further explore trends in adherence to screening over four consecutive rounds in. So far, FIT screening has a relatively low sensitivity for advanced adenomas. The combination of FIT with other non-invasive markers may in theory improve sensitivity without losing specificity. Hence, in **Chapter 7** we will evaluate the use of the gut microbiome in FIT screening as an additional biomarker.

In **PART III**, the value of using the quantative fecal hemoglobin ($_f$Hb) concentration in FIT will be further explored. Currently, FIT is used as a qualitative test even though the precise measurement of fHb in µg Hb allows quantitative results. Therefore, we studied the effect of baseline $_f$Hb concentration in screenees with a FIT below the cut-off (i.e. a $_f$Hb concentration < 10 µg Hb/g feces) on the risk of advanced neoplasia during long-term follow-up in **Chapter 8**. Next, we evaluate the use of different cut-offs for men and women by calculating diagnostic test accuracy over a wide range of fHb concentrations (**Chapter 9**). Screenees with a positive FIT are at high risk of having

advanced neoplasia, often leading to extensive and burdensome colonoscopies. Also, many countries are struggling with colonoscopy capacity and no literature is available on the amount of second-look colonoscopies that come from those high-risk patients. Therefore in **Chapter 10** we tried asses the number of second-look colonoscopies in FIT-based screening and to predict the risk of a second.-look colonoscopy by using fHb. Lastly in **Chapter 11**, we explore the pro's and con's of a hypothetical scenario in which FIT cut-off is lowered so that the screening interval can be extended.

In **PART IV**, quality issues in a CRC screening setting will be discussed. Firstly by assessing the quality of web- based information for screenees, or those contemplating to take part in screening (**Chapter 12**). Both FIT sensitivity and quality of colonoscopy determine the rate of interval carcinomas in CRC screening. In **Chapter 13** the number of interval carcinomas over four rounds of FIT- screening will be evaluated. In addition, screen-detected carcinomas and interval carcinomas will be compared with regard to tumor characteristics, stage and location. To prevent the occurrence of interval carcinomas, as well as preventing unnecessary invasive surgical procedures, recognition of early stage lesions (i.e. T1 lesions) is of paramount importance. This is especially important in screening setting, as more early stage lesions will be detected over the course of multiple rounds. In **Chapter 14** we will investigate the number of early stage lesions detected and adequately diagnosed over four rounds of screening. Moreover, we will focus on the number of unnecessary surgical procedures for these lesions. Lastly, as CRC screening for a great part depends on the execution of the colonoscopy, many quality guidelines and parameters have been established over the past years, and in **Chapter 15** we will compare these quality indicators between both secondary and tertiary hospitals.

In **PART V**, the main findings and conclusions will be summarized and discussed. Finally, future perspectives will be provided.

References

1. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. Gut 2016: [E-pub ahead of print] doi:10.1136/gutjnl-2015-310912.
2. Kuipers EJ, Grady WM, Lieberman D, et al. Colorectal cancer. Nature Reviews Disease Primers 2015: article nmber 15065. doi:10.1038/nrdp.2015.65
3. Zauber AG, Winawer SJ, O'Brien MJ, et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. N Engl J Med 2012; 366(8): 687-96.
4. European Colorectal Cancer Screening Guidelines Working G, von Karsa L, Patnick J, 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.
5. Herrinton LJ, Liu L, Levin TR, Allison JE, Lewis JD, Velayos F. Incidence and Mortality of Colorectal Adenocarcinoma in Persons With Inflammatory Bowel Disease From 1998 to 2010. Gastroenterology 2012; 143(2): 382-9.
6. Schneider EB, Hyder O, Brooke BS, et al. Patient Readmission and Mortality after Colorectal Surgery for Colon Cancer: Impact of Length of Stay Relative to Other Clinical Factors. Journal of the American College of Surgeons 2012; 214(4): 390-8.
7. Winawer SJ, Zauber AG, Ho MN, et al. Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. N Engl J Med 1993; 329(27): 1977-81.
8. Wilson JM, Jungner YG. [Principles and practice of mass screening for disease]
9. Principios y metodos del examen colectivo para identificar enfermedades. Bol Oficina Sanit Panam 1968; 65(4): 281-393.
10. Andermann A, Blancquaert I, Beauchamp S, Dery V. Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. Bull World Health Organ 2008; 86(4): 317-9.
11. Siegel R, DeSantis C, Virgo K, et al. Cancer treatment and survivorship statistics, 2012. CA Cancer J Clin 2012; 62(4): 220-41.
12. Schreuders EH, Ruco A, Rabeneck L, et al. Colorectal cancer screening: a global overview of existing programmes. Gut 2015; 64(10): 1637-49.
13. Kuipers EJ, Rosch T, Bretthauer M. Colorectal cancer screening--optimizing current strategies and new directions. Nat Rev Clin Oncol 2013; 10(3): 130-42.
14. Stoop EM, de Haan MC, de Wijkerslooth TR, et al. Participation and yield of colonoscopy versus non-cathartic CT colonography in population-based screening for colorectal cancer: a randomised controlled trial. Lancet Oncol 2012; 13(1): 55-64.
15. Atkin WS, Edwards R, Kralj-Hans I, et al. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. Lancet 2010; 375(9726): 1624-33.
16. de Haan MC, Pickhardt PJ, Stoker J. CT colonography: accuracy, acceptance, safety and position in organised population screening. Gut 2015; 64(2): 342-50.
17. Bouchard S, Ibrahim M, Van Gossum A. Video capsule endoscopy: perspectives of a revolutionary technique. World J Gastroenterol 2014; 20(46): 17330-44.
18. Rex DK, Adler SN, Aisenberg J, et al. Accuracy of capsule colonoscopy in detecting colorectal polyps in a screening population. Gastroenterology 2015; 148(5): 948-57 e2.
19. Jorgensen OD, Kronborg O, Fenger C. A randomised study of screening for colorectal cancer using faecal occult blood

testing: results after 13 years and seven biennial screening rounds. Gut 2002; 50(1): 29-32.

20. Shaukat A, Mongin SJ, Geisser MS, et al. Long-term mortality after screening for colorectal cancer. N Engl J Med 2013; 369(12): 1106-14.
21. Hol L, Wilschut JA, van Ballegooijen M, et al. Screening for colorectal cancer: random comparison of guaiac and immunochemical faecal occult blood testing at different cut-off levels. Br J Cancer 2009; 100(7): 1103-10.
22. Steele RJ, McClements P, Watling C, et al. Interval cancers in a FOBT-based colorectal cancer population screening programme: implications for stage, gender and tumour site. Gut 2012; 61(4): 576-81.
23. Sung JJ, Ng SC, Chan FK, et al. An updated Asia Pacific Consensus Recommendations on colorectal cancer screening. Gut 2015; 64(1): 121-32.
24. Chiang TH, Chuang SL, Chen SL, 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.
25. Fraser CG, Allison JE, Halloran SP, Young GP, Org WE. A Proposal to Standardize Reporting Units for Fecal Immunochemical Tests for Hemoglobin. Jnci-J Natl Cancer I 2012; 104(11): 810-4.
26. Lee JK, Liles EG, Bent S, Levin TR, Corley DA. Accuracy of fecal immunochemical tests for colorectal cancer: systematic review and meta-analysis. Ann Intern Med 2014; 160(3): 171.
27. Dickinson BT, Kisiel J, Ahlquist DA, Grady WM. Molecular markers for colorectal cancer screening. Gut 2015; 64(9): 1485-94.
28. Jass JR. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. Histopathology 2007; 50(1): 113-30.
29. Leggett B, Whitehall V. Role of the serrated pathway in colorectal cancer pathogenesis. Gastroenterology 2010; 138(6): 2088-100.
30. Imperiale TF, Ransohoff DF, Itzkowitz SH, et al. Multitarget stool DNA testing for colorectal-cancer screening. N Engl J Med 2014; 370(14): 1287-97.
31. Uppara M, Adaba F, Askari A, et al. A systematic review and meta-analysis of the diagnostic accuracy of pyruvate kinase M2 isoenzymatic assay in diagnosing colorectal cancer. World J Surg Oncol 2015; 13: 48.
32. Cipe G, Idiz UO, Firat D, Bektasoglu H. Relationship between intestinal microbiota and colorectal cancer. World J Gastrointest Oncol 2015; 7(10): 233-40.
33. Narayanan V, Peppelenbosch MP, Konstantinov SR. Human fecal microbiome-based biomarkers for colorectal cancer. Cancer Prev Res (Phila) 2014; 7(11): 1108-11.
34. Gagniere J, Raisch J, Veziant J, et al. Gut microbiota imbalance and colorectal cancer. World J Gastroenterol 2016; 22(2): 501-18.
35. zur Hausen H. The search for infectious causes of human cancers: where and why (Nobel lecture). Angew Chem Int Ed Engl 2009; 48(32): 5798-808.
36. Kuipers EJ, de Jong A. Gastrointestinal disorders and Streptococcus bovis bacteremia. Ned Tijdschr Geneeskd 1990; 134(28): 1337-9.
37. Schwabe RF, Jobin C. The microbiome and cancer. Nat Rev Cancer 2013; 13(11): 800-12.
38. Tjalsma H, Boleij A, Marchesi JR, Dutilh BE. A bacterial driver-passenger model for colorectal cancer: beyond the usual suspects. Nat Rev Microbiol 2012; 10(8): 575-82.
39. Di Lena M, Porcelli F, Altomare DF. Volatile Organic Compounds (VOC) as new biomarkers for colorectal cancer: a review. Colorectal Dis 2016.
40. Sonoda H, Kohnoe S, Yamazato T, et al. Colorectal cancer screening with odour material by canine scent detection. Gut

2011; 60(6): 814-9.

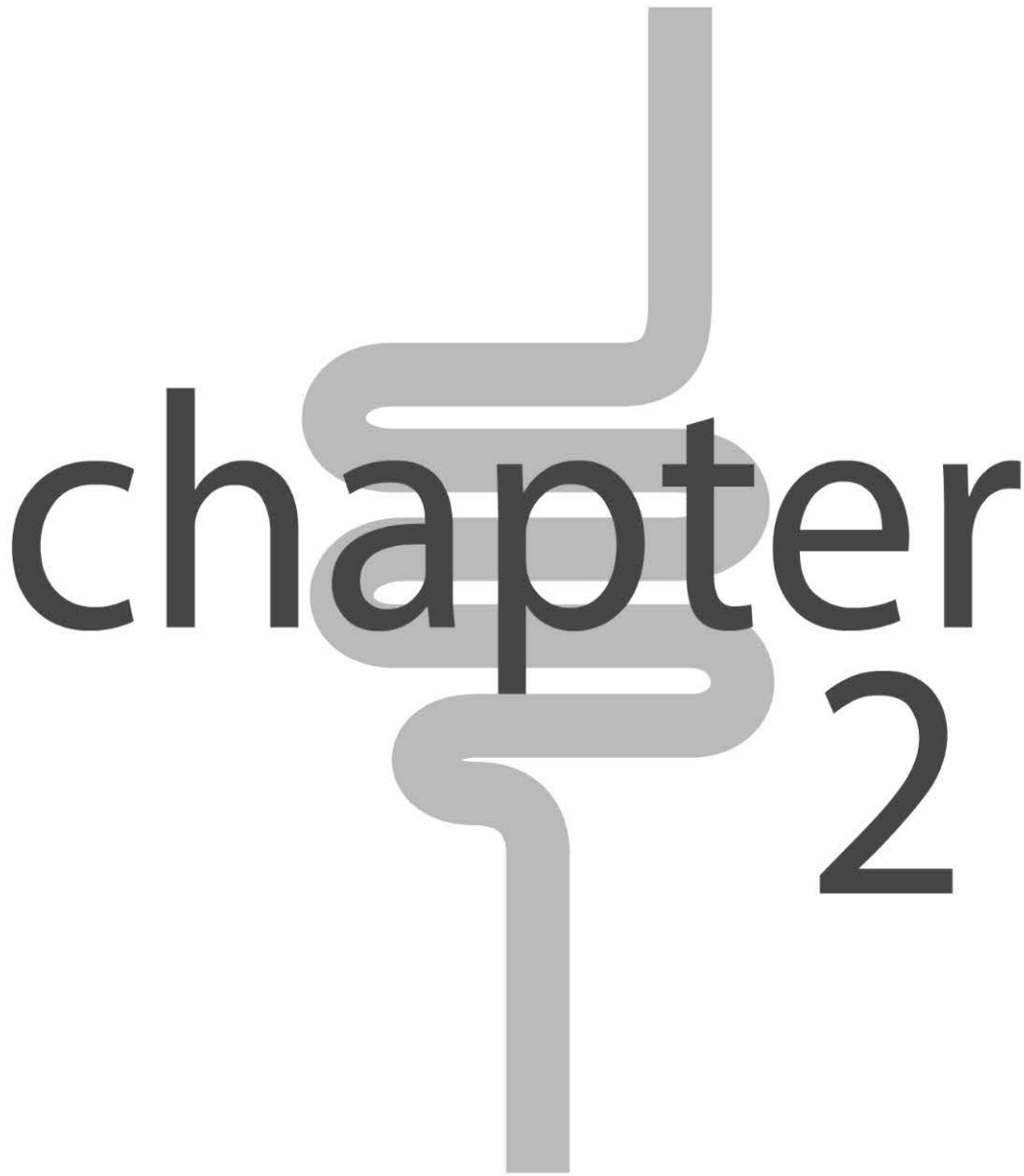
41. Belderbos TD, Grobbee EJ, van Oijen MG, et al. Comparison of cecal intubation and adenoma detection between hospitals can provide incentives to improve quality of colonoscopy. Endoscopy 2015; 47(8): 703-9.
42. de Jonge V, Sint Nicolaas J, Cahen DL, et al. Quality evaluation of colonoscopy reporting and colonoscopy performance in daily clinical practice. Gastrointest Endosc 2012; 75(1): 98-106.
43. Kaminski MF, Regula J, Kraszewska E, et al. Quality indicators for colonoscopy and the risk of interval cancer. N Engl J Med 2010; 362(19): 1795-803.
44. Corley DA, Jensen CD, Marks AR, et al. Adenoma detection rate and risk of colorectal cancer and death. N Engl J Med 2014; 370(14): 1298-306.
45. Rex DK, Schoenfeld PS, Cohen J, et al. Quality indicators for colonoscopy. Gastrointest Endosc 2015; 81(1): 31-53.
46. Atkin WS, Valori R, Kuipers EJ, 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.
47. Aranda-Hernandez J, Hwang J, Kandel G. Seeing better - Evidence based recommendations on optimizing colonoscopy adenoma detection rate. World J Gastroenterol 2016; 22(5): 1767-78.
48. Lai EJ, Calderwood AH, Doros G, Fix OK, Jacobson BC. The Boston bowel preparation scale: a valid and reliable instrument for colonoscopy-oriented research. Gastrointest Endosc 2009; 69(3 Pt 2): 620-5.
49. Radaelli F, Paggi S, Hassan C, et al. Split-dose preparation for colonoscopy increases adenoma detection rate: a randomised controlled trial in an organised screening programme. Gut 2015.
50. Barclay RL, Vicari JJ, Doughty AS, Johanson JF, Greenlaw RL. Colonoscopic withdrawal times and adenoma detection during screening colonoscopy. N Engl J Med 2006; 355(24): 2533-41.
51. Butterly L, Robinson CM, Anderson JC, et al. Serrated and adenomatous polyp detection increases with longer withdrawal time: results from the New Hampshire Colonoscopy Registry. Am J Gastroenterol 2014; 109(3): 417-26.
52. Kaminski MF, Hassan C, Bisschops R, et al. Advanced imaging for detection and differentiation of colorectal neoplasia: European Society of Gastrointestinal Endoscopy (ESGE) Guideline. Endoscopy 2014; 46(5): 435-49.
53. Pohl J, Schneider A, Vogell H, Mayer G, Kaiser G, Ell C. Pancolonic chromoendoscopy with indigo carmine versus standard colonoscopy for detection of neoplastic lesions: a randomised two-centre trial. Gut 2011; 60(4): 485-90.
54. IJspeert JE, Bastiaansen BA, van Leerdam ME, et al. Development and validation of the WASP classification system for optical diagnosis of adenomas, hyperplastic polyps and sessile serrated adenomas/polyps. Gut 2015: [E-pub ahead of print] doi: 10.1136/gutjnl-2014-308411.
55. Paimela H, Malila N, Palva T, Hakulinen T, Vertio H, Jarvinen H. Early detection of colorectal cancer with faecal occult blood test screening. Br J Surg 2010; 97(10): 1567-71.
56. The Paris endoscopic classification of superficial neoplastic lesions: esophagus, stomach, and colon: November 30 to December 1, 2002. Gastrointest Endosc 2003; 58(6 Suppl): S3-43.
57. Endoscopic Classification Review G. Update on the paris classification of superficial neoplastic lesions in the digestive tract. Endoscopy 2005; 37(6): 570-8.
58. Kumar S, Fioritto A, Mitani A, Desai M, Gunaratnam N, Ladabaum U. Optical biopsy of sessile serrated adenomas: do these lesions resemble hyperplastic polyps under narrow-band imaging? Gastrointest Endosc 2013; 78(6): 902-9.
59. Oka S, Tanaka S, Saito Y, et al. Local recurrence after endoscopic resection for large colorectal neoplasia: a multicenter prospective study in Japan. Am J Gastroenterol 2015; 110(5): 697-707.
60. Hewett DG, Kaltenbach T, Sano Y, et al. Validation of a simple classification system for endoscopic diagnosis of small

colorectal polyps using narrow-band imaging. Gastroenterology 2012; 143(3): 599-607 e1.

61. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. Ann Surg Oncol 2010; 17(6): 1471-4.
62. Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 2000; 343(2): 78-85.
63. Giardiello FM, Allen JI, Axilbund JE, et al. Guidelines on genetic evaluation and management of Lynch syndrome: a consensus statement by the US Multi-society Task Force on colorectal cancer. Am J Gastroenterol 2014; 109(8): 1159-79.
64. Schreuders EH, Ruco A, Rabeneck L, et al. Colorectal cancer screening: a global overview of existing programmes. Gut 2015.
65. GLOBOCAN. Estimated cancer incidence, mortality and prevalence worldwide in 2012 http://globocan.iarc.fr/Default.aspx, 2012.
66. van Roon AH, Goede SL, van Ballegooijen M, et al. Random comparison of repeated faecal immunochemical testing at different intervals for population-based colorectal cancer screening. Gut 2013; 62(3): 409-15.

PART II

colorectal cancer screening with a focus on FIT


chapter
2

Guaiac-based fecal occult blood tests versus fecal immunochemical tests for colorectal cancer screening in average-risk individuals

Esmée J. Grobbee, Eline H. Schreuders*, Aafke H.C. van Roon, Leonie van Dam, Ann G. Zauber, I. Lansdorp-Vogelaar, Gerard J.J. M. Borsboom, Ewout W. Steyerberg, Monique E. van Leerdam, Manon C.W. Spaander, Ernst J. Kuipers *both authors contributed equally*

Submitted at Cochrane Diagnostic Test Accuracy Library

Abstract

Background Worldwide, many countries have adopted colorectal cancer (CRC) screening programmes, often based on fecal occult blood tests (FOBTs). These FOBTs fall into two categories based on technique and detected blood component: qualitative guaiac-based FOBTs (gFOBT) and the more recently developed fecal immunochemical tests (FIT) that can be both qualitative and quantitative. Screening by means of gFOBT has been proven to reduce CRC-related mortality. The effectiveness of FIT screening in decreasing CRC-related mortality has not yet been studied in large long-term prospective randomised controlled trials. The primary objective of this review was to compare diagnostic accuracy of gFOBT and FIT screening for detecting advanced colorectal neoplasia in average-risk individuals.

Methods We searched MEDLINE, EMBASE, Cochrane Library, BIOSIS Citation Index, and Science Citation Index Expanded until January 31st 2015. Only studies that provided the number of true positives, false positives, false negatives, and true negatives for gFOBT and/or FIT with colonoscopy as the reference standard were included. Two types of studies were included; those in which all participants underwent both the index test and the reference standard (type I); and those in which only participants with a positive index test underwent the reference standard while the negatives were followed for at least one year for development of interval carcinomas (type II). The target population were asymptomatic, average-risk individuals undergoing colorectal cancer screening were included. Two authors independently selected studies for inclusion and collected the data from each study. In case of doubt a third author made the final decision. We used a bivariate and HSROC statistical model to obtain summary estimates of sensitivity and specificity and summary ROC curves.

Main results A total of twenty-three type I studies involving 85,403 participants were included, reporting on a total of thirty-two fecal occult blood tests. Six studies evaluated gFOBT, thirteen studies evaluated FIT, and four studies included both gFOBT and FIT. Twenty-one studies reported advanced neoplasia as outcome, and nineteen studies reported on colorectal cancer. The cut-off for positivity of FIT varied between 2.4 to 50 μg Hb/g feces. The summary curve estimated by the HSROC model showed that FIT had a higher discriminative ability than gFOBT for advanced neoplasia (p=0.002), and colorectal cancer (p=0.025).

We included nineteen type II studies reporting a total of twenty-three tests involving 1,495,344 participants. Overall six gFOBT studies, ten FIT studies, and three studies combining both gFOBT

and FIT were included. The cut-off for positivity of FIT varied between 2.4 to 10 µg Hb/g feces. The summary curve estimated by the HSROC model showed that FIT had a higher discriminative ability than gFOBT for colorectal cancer ($p<0.001$).

Conclusion FIT is superior to gFOBT in detecting advanced neoplasia and colorectal cancer in average-risk individuals. The specificity of both tests is similar. These results strongly support current guidelines for implementing FIT-based CRC screening programs and the switch from gFOBT to FIT testing for existing programs.

Background

Based on the Wilson & Jungner criteria published in 1968 and updated by the World Health Organization in 2008, screening is justified when (1) a disease is common and associated with significant morbidity or mortality; (2) screening tests are sufficiently accurate in detecting early stage disease, are acceptable to invitees, and are feasible in general clinical practice; (3) treatment after early detection by screening improves prognosis relative to treatment after usual diagnosis; and (4) the potential benefits outweigh the potential harms and costs of screening. Colorectal cancer (CRC) screening fulfils all of these criteria[1,2].

There are various methods for CRC screening. These vary in level of supporting evidence, effectiveness, test-related burden, costs and willingness of target subjects to undergo screening. The screening modalities for CRC broadly fall into two categories; (a) fecal tests (i.e., fecal occult blood tests and fecal DNA testing), and (b) partial or full structural exams (i.e., flexible sigmoidoscopy, colonoscopy and computed tomography-colonography (CTC)). Colonoscopy can be used as the reference standard for those with a positive screening test or as a primary CRC screening tool.
Stool blood tests are conventionally known as fecal occult blood tests (FOBTs), which are used as a two-step testing approach in CRC screening (i.e., positive test result requires further examination with visualization of the colon, predominantly by means of colonoscopy). FOBT screening is based on the principle that a large proportion of colorectal neoplasia bleed microscopically before any clinical signs or symptoms become noticeable. Bleeding tends to be intermittent, and blood is distributed unevenly in the stool. The concept of detecting CRC by testing for blood in the stool is based on the observation that cancers bleed because of disruption of the normal mucosa. The amount of blood increases with the size of the polyp and/or the stage of the cancer[3-6]. In general, the amount of fecal hemeoglobin tends to be absent or low in those without neoplasia, higher for those with advanced adenomas, and highest for those with CRC[7]. Fecal occult blood testing detects a higher proportion of CRCs and a lower proportion of advanced adenomas, since CRCs tend to have a more constant bleeding pattern and give rise to higher amounts of blood in stools than advanced adenomas, which are believed to bleed more intermittently. In this way FOBT screening identifies those individuals who are most likely to have advanced neoplasia. Therefore, it should be followed by visualization of the colon and rectum. Colonoscopy is considered the gold standard for detection of advanced neoplasia with high sensitivity and specificity (both above 90%) and has the advantage that (adenomatous) polyps and early CRCs can be removed during the same procedure. A meta-analysis of the accuracy of colonoscopy (performed for various indications), reported that

the pooled miss rate for adenomas ≥10mm was 2%, for ≥5-10mm 13%, and for 1-5mm 26%[8].
FOBTs fall into two categories based on the detected component of blood: guaiac-based FOBTs (gFOBT) and the more recently developed fecal immunochemical tests (FIT) for hemeoglobin.

Guaiac-based fecal occult blood test

Guaiac-based FOBTs enable detection of occult blood in stool through the pseudo-peroxidase activity of heme. However, peroxidase also reacts with non-human heme present in red meat. Also, several fresh fruits and vegetables contain peroxidase activity, which may lead to false-positive test results. Vitamin C may block the peroxidase reaction, resulting in false-negative test results. Guaiac FOBTs may detect bleeding from any site in the gastro-intestinal (GI) tract, including the stomach, as heme remains relatively stable during transport through the GI-tract[9]. The usual gFOBT protocol consists of three test cards, each containing two panels. The screenee is instructed to collect two fecal samples from three consecutive bowel movements yielding a total of six stool panels. Applying a hydrogen peroxide reagent to the feces on the guaiac material in the panel leads to oxygenation of guaiac, which in turn leads to a blue colour change when heme is present. A panel is considered positive if such coloration appears[10]. The number of positive panels for referral to colonoscopy varies between screening programs. In most programs, a single positive panel is sufficient for referral, however in others the number of positive panels is set at five out of six. In this case, less positive panels imply renewed gFOBT testing. Prior to fecal sampling, individuals are asked to restrict their diet and medication as this might affect the number of false-positive and false-negative test results.

The sensitivity and specificity of gFOBT screening varies widely due to the variation in type of test (brand), instructions for stool collection, number of stool samples per screening round, the use of non-hydrated or rehydrated stool samples, double reading of the test, the number of positive panels used to refer a screened person for colonoscopy, and the interval between successive screening rounds. In some trials, rehydrated gFOBT has been studied; rehydration reduces the false negative rate (improves sensitivity) while increasing the false positive rate (reduces specificity)[11,12].
Guaiac FOBTs are the only stool tests for which there is evidence of efficacy from four prospective, randomised controlled trials (RCTs). These trials from the USA, United Kingdom, Denmark and Sweden demonstrated that multiple rounds of annual or biennial gFOBT screening can reduce CRC-related mortality by approximately 13-33%[11,13-15]. The American trial, which used rehydrated gFOBT, also demonstrated a reduction in the incidence of CRC[11]. A subsequent meta-analysis reported a pooled 15% reduction in CRC-related mortality among the three biennial screening trials with gFOBT compared to controls16. The American trial recently reported an overall reduction in CRC mortality

of 27% after 30 years of follow-up[17]. The efficacy of gFOBT screening in reducing CRC-related mortality is limited due to a limited sensitivity for detecting CRC and low sensitivity for detection of advanced adenomas[18]. Furthermore, the process of analysing gFOBTs is time consuming and is faulted by the possibility of inaccurate processing and evaluation[19].

Fecal immunochemical test

FITs have several technological advantages compared to guaiac based screening. FIT specifically targets human globin, a protein that along with heme constitutes the hemeoglobin molecule. Therefore, FITs only detect human blood, in contrast to the gFOBT which can falsely detect other substances. For this reason, FITs are less subject to interference by dietary factors and medication. Studies have suggested that NSAID or aspirin use increased the sensitivity of FIT without a decrease in specificity[20,21]. In addition, FITs are more specific for lower GI-tract bleeding since globin is degraded by digestive enzymes in the upper GI-tract. This improves their specificity for neoplasia in the colon and rectum. The sample collection for most FIT variants is less demanding than for gFOBT-sampling, both in terms of requiring a single sample and less direct handling of stools (smear cards for gFOBTs vs brush/spatula for FIT testing). Furthermore, FIT screening does not require dietary restrictions. Both qualitative and quantitative FITs have been developed and are described below.

Qualitative FITs

Qualitative tests require a manual interpretation of test results as positive or negative. There is a range of such tests on the market. They often use immunochromatographic technology, and allow for simple, office-based analysis. Since qualitative FITs provide dichotomous test results and thresholds for a positive test differ between brands, test performances differ[22,23]. However, like gFOBT, inter-observer variations in interpretation of test results may influence performance.

Quantitative FITs

Quantitative FITs on the contrary can be analysed automatically, quantifying the amount of hemeoglobin found in the stool sample. One advantage of quantitative FITs in CRC screening programs is that the cut-off level (i.e. the amount of hemeoglobin above which the test is considered positive and individuals are referred for follow-up examination) can be adjusted. This allows the number of FIT-positives to be matched with the available resources for further investigation, in particular colonoscopy capacity[24].Quantitative FITs have further important advantages over qualitative FITs due to the use of automated analysis. This automation removes inter-observer variation in interpretation of test results, improves reproducibility, and allows for high-throughput

testing. Nevertheless, studies suggest variable performance of different brands of quantitative FITs, even when the standardized same cut-off is used[22].To date, there are no long-term prospective randomized data that demonstrate that FIT is superior to gFOBT in terms of reducing CRC-related mortality. However, a recent ecological study compared regions in Italy with and without population FIT screening. CRC-specific mortality was 22% lower in areas with a FIT screening program compared with areas without a screening program[25].With this review, we aim to compare the diagnostic test accuracy measures of gFOBT and FIT screening in order to answer the question "Can gFOBT be replaced by FIT for primary CRC screening?". In order to answer this question, an overview of the test performance characteristics for both types of FOBT will be provided.

Target condition being diagnosed

FOBT screening primarily aims at early detection of bleeding colorectal neoplasia, since only bleeding lesions can be detected by stool blood tests. CRC screening in general aims at lowering CRC mortality by early detection of CRC and lowering CRC incidence by removal his pre-malignant lesion i.e., adenomatous polyps.

Index tests

The tests under evaluation are two FOBTs: gFOBT and FIT. More detailed information about the tests and the methods of execution have been previously described. FIT can be both quantitative as well as qualitative, the latter does not report individual fecal Hb concentrations.

Alternative tests

There are several alternative tests that can be used for CRC screening purposes. These tests vary in the level of supporting evidence, attendance, effectiveness, and test-related burden, costs. Alternative screening modalities usually considered as effective CRC screening tools include flexible sigmoidoscopy, colonoscopy, computed tomography-colonography and, more recently, capsule endoscopy, fecal DNA testing and serum molecular markers.26,27

Rationale

In the Western world, many countries have adopted a CRC screening program, often based on FOBT.28 Screening by means of gFOBT has been proven to reduce CRC-related mortality. The results on effectiveness of FIT screening in decreasing CRC-related mortality are not yet available. The main explanation for this is that many countries have already implemented a CRC screening program. In addition, decisions on the optimal screening test have to be based on data about the sensitivity and

specificity, existing RCT results, and modelling[29].

Objectives

The primary objective of this review is to compare the diagnostic test accuracy of gFOBT and FIT screening for detecting advanced colorectal neoplasia in average-risk individuals.

Investigation of sources of heterogeneity

We aimed to investigate the following sources of heterogeneity.

A. Heterogeneity related to characteristics of the study population (i.e. sex, age limits, ethnicity, selection of invitees (identified from general practitioner records or population registers), cancer stage, distribution and cancer location.

B. Heterogeneity related to the number of FOBTs performed per screening round

C. Heterogeneity related to the cut-off value used for FIT or the number of positive panels used to refer a gFOBT-screened person for colonoscopy.

Due to reasons described later in this review, analysis for heterogeneity could not be performed for all of these factors.

Methods

Criteria for considering studies for this review

Types of studies: Two different types of studies were included and categorized in this review:

Type I studies: All (randomised, comparative) accuracy studies in which all participants underwent both the index test and the reference standard. Diagnostic case-control studies were considered inappropriate for this review because such studies are likely to overestimate diagnostic performance[30]. Moreover, literature suggests that measures of accuracy may vary with the prevalence and stage-distribution of the target condition[31].For instance, the sensitivity of a test will often vary according to the severity of the detected disease (e.g. advanced CRCs are more easily detected with FOBTs than early stage tumours). For these reasons, we did not include case-control studies in this review.

Type II studies: All (randomised, comparative) accuracy studies in which all participants with a positive index test were referred for the reference standard and all participants with a negative index test were followed for at least one year to identify development of interval carcinomas. Only data from the first screening round were included for analysis.

Participants

Asymptomatic average-risk individuals aged 40 years and above were considered as representative for a CRC screening program. Study participants included subjects volunteering for a medical health check-up (including CRC screening), as well as individuals identified from population registers, and general practitioner or managed care organisation records.

Index tests

The index test was either gFOBT or FIT (both qualitative and quantitative) as described previously in the Background section.

Comparator tests

Studies were included regardless of whether they made comparisons with other CRC screening modalities.

Target conditions

The primary target condition was CRC, which was defined as the invasion of malignant cells beyond the lamina muscularis mucosa. Patients with an intra-mucosal carcinoma or carcinoma in situ were classified as having high-grade dysplasia[32].The secondary target condition was advanced neoplasia, which included CRC and advanced adenomas. An advanced adenoma was defined as an adenoma with a greatest dimension of at least 10 mm, or an adenoma with ≥ 25% villous component, and/or high-grade dysplasia[32]. For each included study, we assessed whether these definitions were applied. If another definition was adopted in a study, we stated this in the characteristics of included studies.

Reference standards

Studies were included for this review if colonoscopy was used as the primary reference standard. Only in case of an incomplete colonoscopy, CTC or double contrast barium enema (DCBE) was accepted as reference. Furthermore, in type II studies participants with a negative index test result had to be followed for at least one year to assess the development of interval carcinomas. Interval carcinomas were defined as CRC diagnosed in an FOBT negative screenee in the period between two successive FOBT screening rounds[33].If a study did not use this definition, we stated this in the characteristics of included studies.

Exclusion criteria

We excluded studies where more than 5% of the population consisted of high-risk individuals. High-risk individuals were defined as patients with a history of CRC; subjects with a personal history of adenoma(s); individuals scheduled for diagnostic colonoscopy because of hereditary CRC syndromes or a positive family history of CRC; symptomatic subjects with complaints suspicious for CRC such as rectal blood loss, changed bowel habits, or weight loss; and all patients with a history of inflammatory bowel disease. We also excluded studies in which a positive gFOBT test result needed to be confirmed by a positive FIT test result or vice versa. We excluded studies in which less than 75% of the participants with a positive FOBT underwent colonoscopy or in case of an incomplete colonoscopy, CTC or DCBE.

Electronic searches

To identify appropriate studies, the Trial Search Coordinator of the Cochrane Colorectal Cancer Group in collaboration with the Medical School Library of the Erasmus MC conducted a literature search by using the electronic databases Medline, Embase, the Cochrane Library, BIOSIS and SCI-expanded. The Embase, Medline and Biosis searches were run in OVID. There were no restrictions on date or language of the articles being reviewed. Native speakers related to our departments and personal acquaintances translated articles written in languages other than English. The searches were developed using the Boolean term 'AND' between the topics colorectal cancer and fecal occult blood test. To cover the topic of colorectal cancer, the searches were developed by searching MeSH and/or EmTree terms colorectal neoplasms, colorectal cancer and large intestine tumour, and the text words: colorectal, rectal, rectum, colon*, cancer*, carcinoma*, adenocarcinom*, neoplas*, tumor*, tumour*, polyp* and adenom*. To cover the topic of fecal occult blood testing, the search was developed by searching MeSH and/or EmTree terms occult blood, immunochemistry and feces analysis, and the text words; fecal, fecal, feces, feces, stool*, occult blood, occult blood test*, FOBT*, gFOBT*, FIT*, immunochem*, immunological*, guaiac*, fecal immunochem*, fecal immunochem* and test*. In exploratory searches, we identified articles that used gFOBT brand names without explicitly mentioning either gFOBT or FOBT in the title, abstract or MeSH terms. We have therefore incorporated most of the gFOBT brand names in our literature search. This brand name issue was not present when searching for articles related to the FIT. Differing from the previous published search strategy in our protocol, we did not include the brand name "colorectal" as this yielded too many irrelevant results. Initial searches were conducted in September 2013. A second search was conducted on January 31st, 2015. We performed a further search on March 21st, 2016. Those results have been added to 'Studies awaiting classification' and will be incorporated into the review at the

next update.

Searching other resources

The references of all included relevant studies were hand-searched for additional trials. In addition, we searched for articles citing the relevant studies included in the review. We defined relevant included articles as any included article that was published within the 5 years preceding our search. Furthermore, we searched PubMed for Related Articles of the most relevant included articles. We examined the first 20 results from 'PubMed Related Articles' after sorting by publication date from newest to oldest. We also contacted principal investigators of the included articles to clarify aspects of methods and results, and ask for any unpublished data in the area of FOBT characteristics, where necessary.

Selection of studies

Two reviewers (EJG and EHS) independently assessed whether the titles and abstracts were eligible for further reading. After this initial retrieval, all selected articles were read entirely. Disagreements about including a study for this review were resolved through discussion with a third reviewer (AvR). All studies that did not meet the inclusion criteria, as ascertained in reading the full article, were listed in a separate table with reasons for exclusion. The reference management software EndNote X7 was used for the selection process.

Data extraction and management

Data were extracted from those trials that fulfilled the inclusion criteria. The data that were extracted for both advanced neoplasia and CRC were:

- Positivity rate (PR), i.e. the proportion of participants having a positive index test result.
- True positives (TP), i.e. participants having a positive index test result, followed by detection of advanced neoplasia by means of the reference standard.
- False positives (FP), i.e. participants having a positive index test result, but no advanced neoplasia when assessed with the reference standard.
- True negatives (TN), i.e. participants having a negative index test result, and no advanced neoplasia during colonoscopy for Type I studies and no interval CRC identified during follow up for Type II studies.
- False negatives (FN), i.e. participants having a negative index test result, and advanced neoplasia during colonoscopy for Type I studies and interval CRC identified during follow up for Type II studies.

The analyses only include the main outcome measures sensitivity and specificity (which were derived

from TP, FP, TN, FN). For CRC we included data from both type I and II studies, and for advanced neoplasia we included data from type I studies only. For both study types, the extracted data were merged into separate 2 x 2 tables (containing TP, FP, FN, TN). We excluded non-interpretable test results and FOBT-positives who refused to undergo the reference standard from the 2x2 table and in consequence from the meta-analysis. If data were lacking in a specific article we contacted the principal investigators to ask for the original data and/or tried to reconstruct the aforementioned cell frequencies from the information that was published. If this was not successful, we excluded the study. The data presented in the 2 x 2 table were used to conduct meta-analysis on sensitivity and specificity. For type II studies, only data regarding CRCs were generally available during follow-up. Therefore, for type I studies we were able to extract data for both advanced adenomas and CRC, but for type II studies we extracted data for CRC only. We extracted data for all possible cut-offs. Since the concentration used for cut-off [ng Hb/ml buffer in the device] is unique to the device or system and cannot be compared with other devices, cut-offs were transformed to the internationally accepted unit of µg Hb/g feces[34]. All data were extracted independently by two reviewers (EJG and EHS).

Assessment of methodological quality

Two authors (EJG and EHS) independently assessed the quality of each individual study using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool[35]. We excluded some of the questions from the QUADAS-2 tool in case they were not applicable. Details of each study are described in the Characteristics of Included studies table. If data was not specified in the article, this was mentioned. When authors did not respond also manufacturers were contacted to retrieve additional details about the test used if needed.

Descriptive analysis

The descriptive analysis provides an overview of all available studies. Tables were split by gFOBT or FIT, and by type I or type II studies. For all study types, the following test characteristics were extracted into 2x2 tables: TP, FP, TN, and FN. The extracted data were entered into Review Manager 5. Study-specific estimates and exact 95% confidence intervals (CI) of sensitivity and specificity were obtained and displayed in forest plots per test type. Different symbols were used per test type, in order to create a clear overview of between-test variability.

Inferential statistics

In secondary analyses, we compared the performance of the gFOBTs and versus the FITs. We complied with the methods and techniques introduced and explained in chapter 10 of the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy[36].We started with an exploration of the study-specific sensitivities and specificities that were extracted from the included studies using RevMan software. Based on the available reported sensitivities and specificities, we used the bivariate model, and the Rutter and Gatsonis Hierarchical summary receiver-operator curves (HSROC) model to explore differences between tests and identify potential sources of heterogeneity[37]. When at least three studies were found in which both gFOBT and FIT were compared with the reference standard we analysed these as a separate group from studies in which only one of the two tests was compared. When less of these studies were found, we allowed that these contributed an observation to both series of tests. These studies were then included twice in the data set. We also calculated initial summary estimates of sensitivity and specificity when it turned out that more than three studies using common cut-off points for test positivity were available. If enough studies appeared per test brand with the same outcome parameters, we performed meta-analyses. The HSROC model was used to analyse sensitivities and specificities for type I and type II studies separately, including all studies. For the quantitative FIT where 2x2 tables for multiple cut-offs were available, we used the cut-off as advised by the manufacturers for this analyses.

We investigated the effect of cut-off by carrying out subgroup meta-analyses for cut-offs where sufficient data were available. As some studies reported 2 x 2 data for more than one cut-off, this analysis allowed us to include all of the available data. To analyse the sensitivities and specificities from the various studies and test types that used the same cut-off, we used the bivariate model as introduced by van Houwelingen et al., extended by Reitsma et. al., and explained in the Cochrane Manual Chapter 10[36,38,39]. The bivariate model was fitted using proc nlmixed from SAS version 9.2 conforming to the examples of SAS syntax given in chapter 10 of the Manual. We primarily compared the accuracy of gFOBT and FIT tests, employing a model that had the same between study variance/ covariance matrix in these two test types. In order to include studies that had zero counts in any of the four cells we added 0.5 to cells containing no observations. Model output provided confidence and prediction region parameters and summary estimates of test accuracy measures per test type (gFOBT or FIT).

Investigations of heterogeneity

The I2 statistic40 was not calculated as it doesn't account for heterogeneity explained by phenomena

such as positivity threshold effects, it is therefore not routinely used in Cochrane DTA reviews. The magnitude of observed heterogeneity was depicted graphically by the prediction ellipse. We planned to address heterogeneity by adding covariates of interest to the HSROC model. The factors that we aimed to include in our heterogeneity analyses are described in our objectives. However, there are several caveats to keep in mind:

A. Heterogeneity related to gender was assessed as percentage of male participants; investigation of heterogeneities within other population characteristics was not feasible due to lack of information provided in individual studies. Nevertheless, only (studies with) average-risk individuals, as defined in our protocol, were included. The well-defined criteria resulted in homogenous studies according to prediction ellipses. Investigation of heterogeneities within cancer stage, distribution and cancer location was not feasible due to lack of information provided in individual studies.

B. Heterogeneity related to the number of tests and/or the number of stools per screening round was assessed.

C. Summary estimates of sensitivity and specificity for the two most common cut-off levels to refer screenees for further evaluation by the reference standard (i.e. cut-off value for FIT in µg/ml or the number of positive panels for gFOBT) were performed. Heterogeneity related to the quantitative or qualitative nature of FIT was assessed.

Sensitivity analyses

Sensitivity analyses were performed in which the QUADAS-items were used to identify studies that scored differently on certain quality items to determine the effect of poor study quality on the overall results. The impact of each study was tested by removing each one from the analysis separately and recalculating the summary estimates.

Assessment of reporting bias

Investigation of publication bias in diagnostic test accuracy studies has proven to be problematic, because many studies are done without ethical approval or study registration[41-43]. Therefore, identification of studies from registration until final publication of the results is not possible[41]. Furthermore, funnel plot-based tests that are commonly used to detect publication bias in reviews of randomised controlled trials, have been shown to be misleading for diagnostic test accuracy reviews such as ours. Therefore we did not assess reporting bias.

Results

Results of the search

The search identified 5,355 titles, of which 2,797 remained after removal of duplicates. Of these, 2,194 were excluded on the basis of title and abstract. Manual removal of duplicates for the remaining articles resulted in 38 additional duplicate articles. From 40 articles, the PDF could not be retrieved, even after trying to contact the authors. Full articles were retrieved for 369 titles. After hand-searching the references of all included articles and PubMed related articles of included studies, nine additional articles were identified and fully assessed. A total of 378 full-text articles were assessed for eligibility, of which 337 articles were excluded because they met one of the exclusion criteria or were otherwise assessed as ineligible for the following reasons:

In 162 studies, only FOBT-positive subjects had undergone the reference standard without follow-up of FOBT-negatives:

- Forty-seven studies did not focus on average-risk subjects;
- Thirty-six studies had not used colonoscopy as the first choice of reference standard;
- Fifteen studies only provided data on cumulative mortality over multiple screening rounds making it impossible to determine absolute numbers of advanced neoplasia detected per screening round;
- Thirteen studies the full-text was a letter or editorial, and thirteen were reviews;
- Fourteen articles encompassed the same cohort as an already included article;
- Six articles summarized the results of multiple screening rounds and separate data-extraction of the first round was not possible;
- Four articles focussed on digital FOBT where a stool sample was obtained by digital rectal examination;
- Twenty-seven articles were excluded for various other reasons (Figure 1).

A total of 41 studies were included. One additional study (Faivre 2004) was included after contacting the author to obtain data on another article to be included[44]. We combined two articles (Brenner 2013, Haug 2011) for gFOBT and FIT results[45,46].Both studies analysed the same population of 3,077 patients. For the Haug's study the authors provided results for different FIT cut-offs to allow direct comparison with other studies using the same cut-off. The 15 excluded cases in the original article for the analysis about left/right sided lesions were included in our analysis after contacting the principal investigators.

Table 1. *Overview test characteristics per study for type I studies.*

Study	Test brand	gFOBT	FIT	FIT method	FIT 10 µg	FIT 20 µg	AN	CRC	Other cut-off	Nr of stools
Alhquist 2008 [47]	Hemocult	+					+	+		3
	Hemocult Sensa	+					+	+		3
Brenner G 2010 [48]	HemoCARE (gFOBT)	+					+	+		3
	ImmoCARE-C (FIT)		+	Qualitative			+	+	unknown	3
Brenner H 2013 [45]	Hemoccult	+					+	+		1
Chen 2014 [49]	OC-light		+	Qualitative	+		+	+		1
Cheng 2002 [50]	OC Hemodia		+	Qualitative	+		+	+		1
Chiu 2013 [51]	OC-light		+	Qualitative	+		+	+		1
Cruz-Corrrea 2007 [52]	Hemocult II	+					+	+		3
de Wijkerslooth 2012 [53]	OC-sensor		+	Quantitative	+	+	+	+		1
Graser 2009 [54]	brand not specified	+					+			3
	FOB-Gold		+	Quantitative			+	+	14 ng/ml= 2.4 µg/g	2
Haug 2011 [46]	RIDASCREEN		+	Quantitative	+	+	+	+		1
Hernandez 2014 [55]	OC-sensor		+	Quantitative	+	+	+	+		1
Hoepffner 2006 [56]	Hemocult	+					+	+		1
	Hb ELISA Immunodiagnostik		+	Unknown	+		+	+	10 µg/g	1
Imperiale 2004 [57]	Hemocult II	+					+			3
Imperiale 2014 [58]	OC-sensor		+	Quantitative		+	+			1
Khalid – de Bakker 2011 [59]	OC-sensor		+	Quantitative	+	+	+	+		1
Levy 2014 [60]	Inverness Clearview		+	Qualitative			+	+	50 µg/g	1
	Alere Clearview		+	Qualitative			+		6 µg/g	1
	Polymedco OC-Light		+	Qualitative	+		+	+		1
	Quidel QuickVue		+	Qualitative			+	+	50 µg/g	1
Lieberman 2001 [12]	Hemocult II	+					+	+		3
							+	+		

Nakama 2000 [61]	Iatro Hemcheck		+	Qualitative				+	unknown	2
Omata 2011 [62]	OC-micro		+	Quantitative	+	+	+	+		1
Park 2010 [63]	Hemocult II	+					+	+		3
	OC-sensa		+	Quantitative	+	+	+	+		3
Sung 2003 [64]	Hemocult II	+					+	+		3
Wong 2014 [65]	Hemosure		+	Qualitative			+	+	50ng/ml = 50 µg/g	1
Wu 2014 [66]	ACON Laboratories		+	Qualitative			+	+	50ng/ml = 6 µg/g	1

Table 2. *Overview test characteristics per study for type II studies.*

Study	Test brand	gFOBT	FIT	FIT method	FIT 10 µg	FIT 20 µg	CRC	Other cut-off	Nr of stools
Bouvier 1999 [67]	Hemocullt II	+							n.d.
Castiglione 2007 [68]	OC-Hemodia		+	Quantitative		+			1
Chiang 2014 [22]	OC-sensor		+	Quantitative		+			1
	HM-Jack			Quantitative		+			1
Crotta 2012 [69]	OC-sensor		+	Quantitative		+			1
Denters 2012 [70]	Hemoccult II	+							3
	OC-Sensor		+	Quantitative	+				1
Faivre 2004 [44]	Hemocullt II	+							3
Giai 2014 [71]	Hemocullt II	+							3
Itoh 1996 [72]	OC-Hemodia		+	Qualitative	+				1
Kronborg 1987 [73]	Hemocullt II	+							3
Launoy 2005 [74]	Magstream		+	Quantitative				20 ng/ml= 67 µg/g	2
Levi 2011 [75]	Hemocullt Sensa	+							3
	OC-micro		+	Quantitative				70 ng/ml= 14 µ/g	3
McNamara 2014 [77]	OC-sensor		+	Quantitative		+			2
Paimela 2010 [78]	Hemocullt	+							3
Parra-Blanco 2010 [79]	Hemofec	+							3
	OC-light		+	Qualitative	+				1
Parente 2013 [80]	HM-Jack		+	Quantitative				100 ng/ml= 250 µg/g	1
Steele 2009 [81]	Hema-screen	+							3
Sieg 2002 [82]	unknown		+	Quantitative				5 µg/g and 10 µg/g	1
Van Roon 2013 [83]	OC-Sensor		+	Quantitative	+				1

A total of 23 type I studies were included: five were performed in the United States, five in Germany, four in Taiwan, two in the Netherlands, two in Japan, and the remaining studies were performed in France, Spain, South-Korea, China and Hong Kong. Seven studies compared more than one test and of those there were six studies in which participants had undergone more than one index test; resulting in 91,971 test evaluations with a total of 32 separate tests in 85,403 participants. Overall, six gFOBT studies, thirteen FIT studies, and four studies combining both gFOBT and FIT screening were included for this review. The earliest study was published in 2000 (Nakama 2000), with the majority being published between 2008 and 2013. For all but one study (Nakama 2000) advanced neoplasia was the main outcome. Twenty papers separately described the numbers of detected CRC and advanced adenoma. In two studies, no CRC was detected. For eleven studies, data for a cut-off of 10 µg Hb/g feces could be retrieved, and for eight studies a cut-off of 20 µg Hb/g feces was used. For all included gFOBT studies a positivity criterion of at least one positive panel was used. All but two studies used a single stool sample for FIT-testing and for all but two gFOBT-studies three consecutive stools were used (Table 1).

A total of nineteen type II studies were included: four were performed in France, three in Italy, two in Japan, two in the Netherlands, and the remaining studies were performed in Taiwan, Denmark, Isrel, Ireland, Scotland, Finland, Spain (Tenerife), and Germany. Four studies compared more than one index test with each participant undergoing one FOBT. In total they reported on a total of 23 tests in 1,495,344 participants. Overall, six gFOBT, ten FIT, and three combination studies were included. Out of the studies combining gFOBT and FIT, two studies randomized participants and one study performed both tests in all participants. The earliest study was published in 1987 (Kronborg 1987), with the majority being published in 2002 to 2014. All studies had at least one year of follow-up, with a maximum of four years of follow-up. Eleven of nineteen included studies had exact two years of follow-up. For four studies, data with a cut-off of 10 µg Hb/g feces could be retrieved, and for four studies (five FITs) data with a cut-off of 20 µg Hb/g feces. All but one study used a single stool sample for FIT-testing and for all gFOBT-tests, three consecutive stools were used (Table 2).
An updated search in March 2016 identified 391 additional records. A total of 355 records were excluded based on title and/or abstract, 36 study reports from this updated search have been added to 'Studies awaiting classification'.

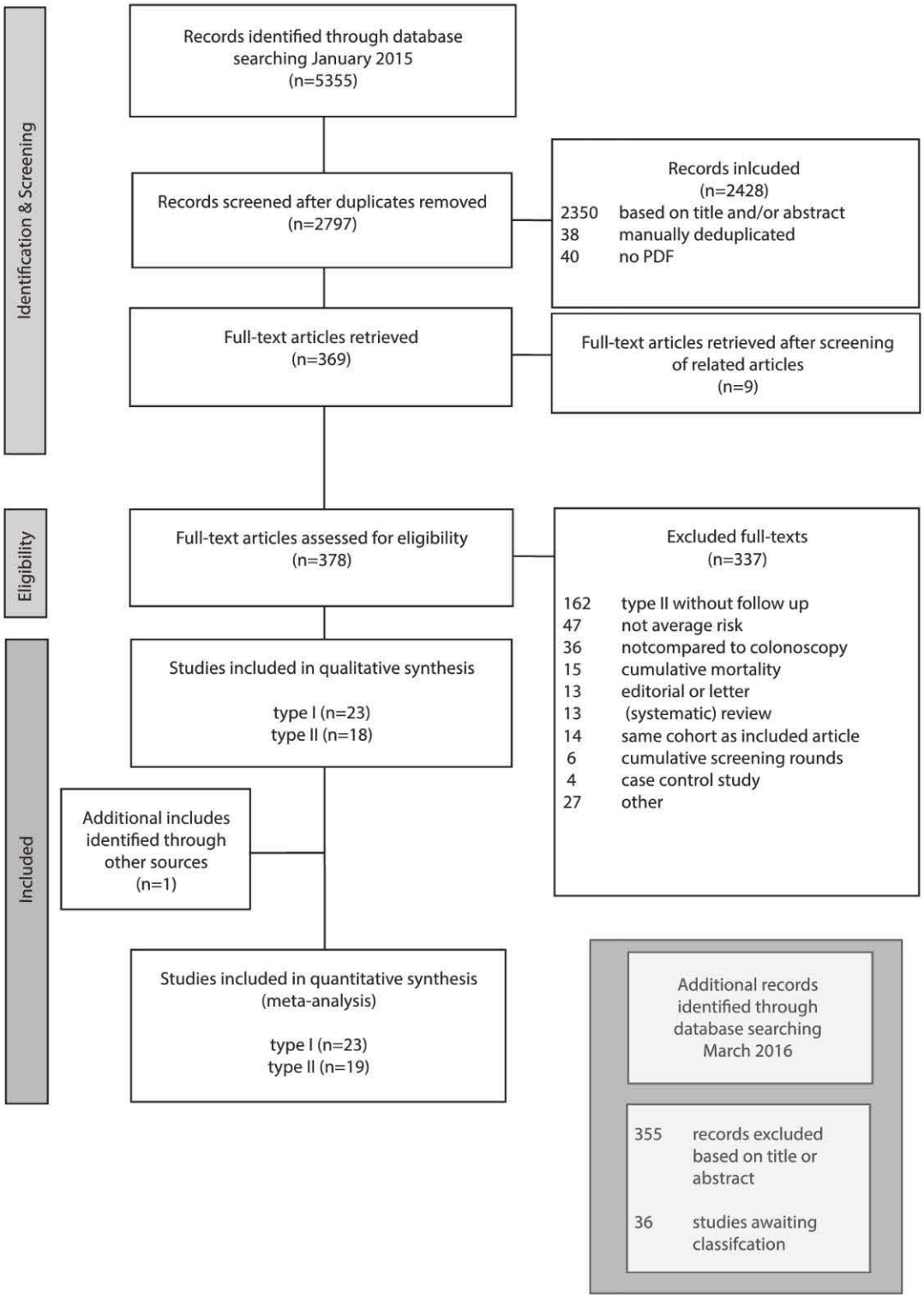


Figure 1. *Flowchart of search and included studies.*

Methodological quality of type I studies

The overall quality of type I studies is summarized in Figures 2A and 2B using QUADAS-II. Nineteen (83%) studies clearly included a representative spectrum of average-risk participants, which reflects population-based screening. For three studies this spectrum was unclear because the studies it were either retrospective or did not describe exclusion criteria clearly. However, all these studies were performed in an average-risk CRC screening setting and therefore included for analysis. One study (Cruz-Correa 2007) had a high risk of bias regarding the spectrum as it included patients referred for colonoscopy outside of a screening setting. However, only asymptomatic patients over the age of 55 years were included, for these reasons we choose to include this article for analysis. Eighteen (78%) of the studies had a low risk of bias concerning the index test; four were unclear because either the method of collection was not clearly described or the positivity threshold was not described. One study (4%) had a high risk of bias since the study conducted the index test differently than as advised by the manufacturer. Unanalysable tests were only reported in five studies (22%). Twenty-two studies (96%) had low concerns regarding applicability of the index test; with one study rated as unclear because this study did not describe whether the threshold was pre-specified. All studies had a low risk concerning the reference standard with over 80% of the participants undergoing colonoscopy as the reference standard. However, many studies had missing values and FOBT-positivity rates were often lacking. The majority (87%) of studies had clear definitions of advanced adenomas; mainly defined as adenomas ≥10 mm, adenomas with at least 25% villous component, and /or high grade dysplasia.

Methodological quality of type II studies

The overall quality of type II studies is summarized in Figures 3A and 3B. Fourteen studies (74 %) clearly included a representative spectrum of participants with an average-risk of developing advanced neoplasia. Two studies had a high risk of bias with regard to selection of patients; in one study (Itoh 1996) Japanese workers in a FIT-based screening programme could have experienced gFOBT screening during earlier years. In another study (Sieg 2002), the article stated that subjects below the age of 44 could also participate if they heard of the study, but when contacted the authors stated this was not the case. Risk of bias concerning the index test was potentially present if the paper did not specify how the authors had handled non-interpretable or borderline test results. Around 84% specified their reference standard as being colonoscopy. Three studies were marked as unclear because while they used colonoscopy as a reference standard, they did not describe how many people underwent CT-colonography or DBCE in case of a failed colonoscopy. With regard to flow and timing, 47% of the studies had a low risk of bias. This was due to multiple reasons. All

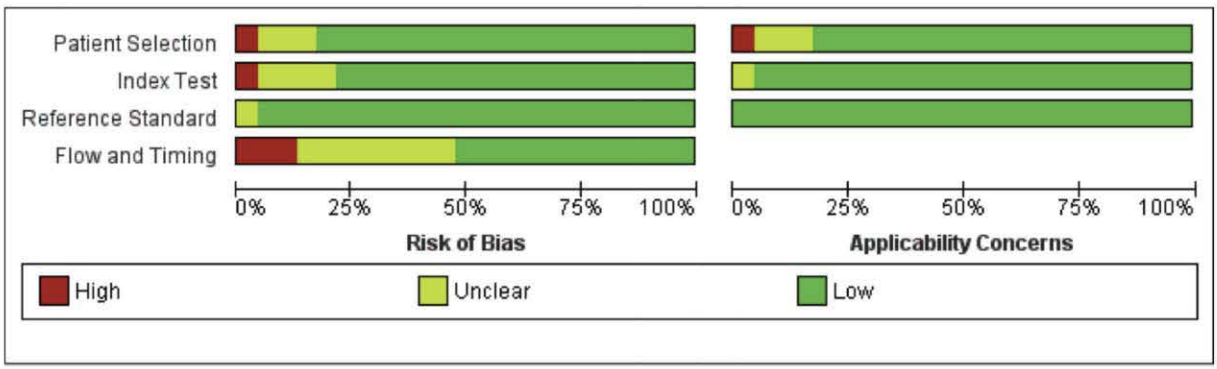


Figure 2A. Type I studies. *Risk of bias and applicability concerns graph: review authors' judgements about each Fdomain presented as percentages across included studies*

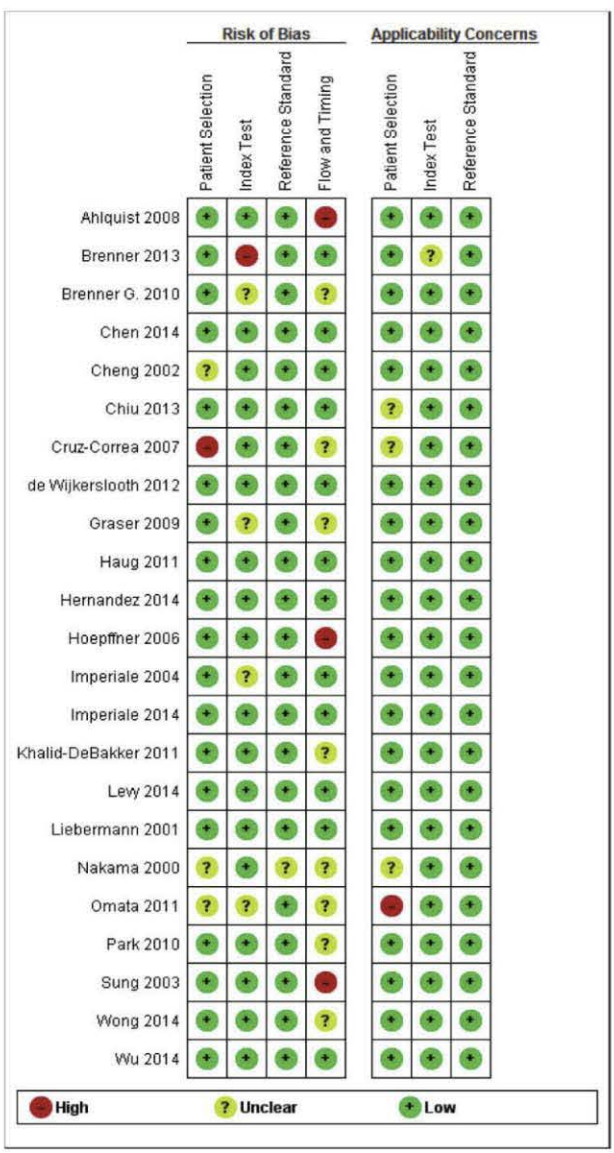


Figure 2b. Type I studies. *Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study*

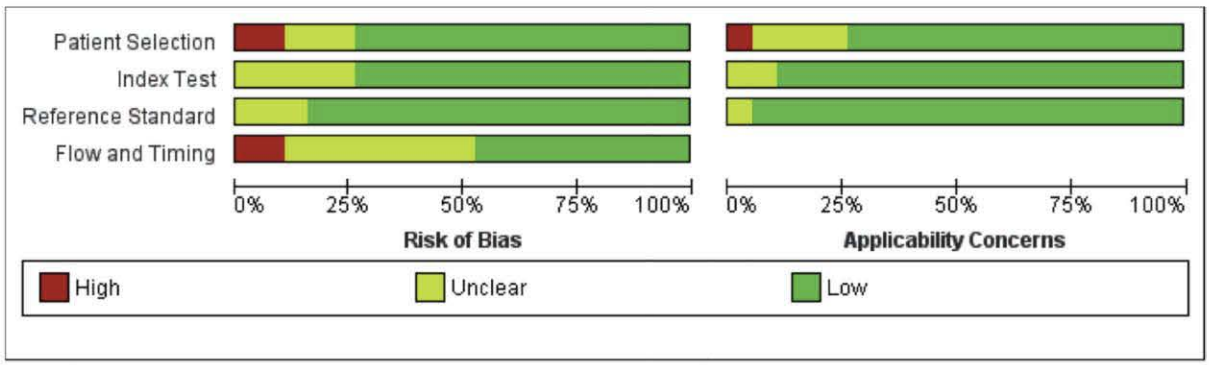


Figure 3a. Type II studies. *Risk of bias and applicability concerns graph: review authors' judgements about each domain presented as percentages across included studies*

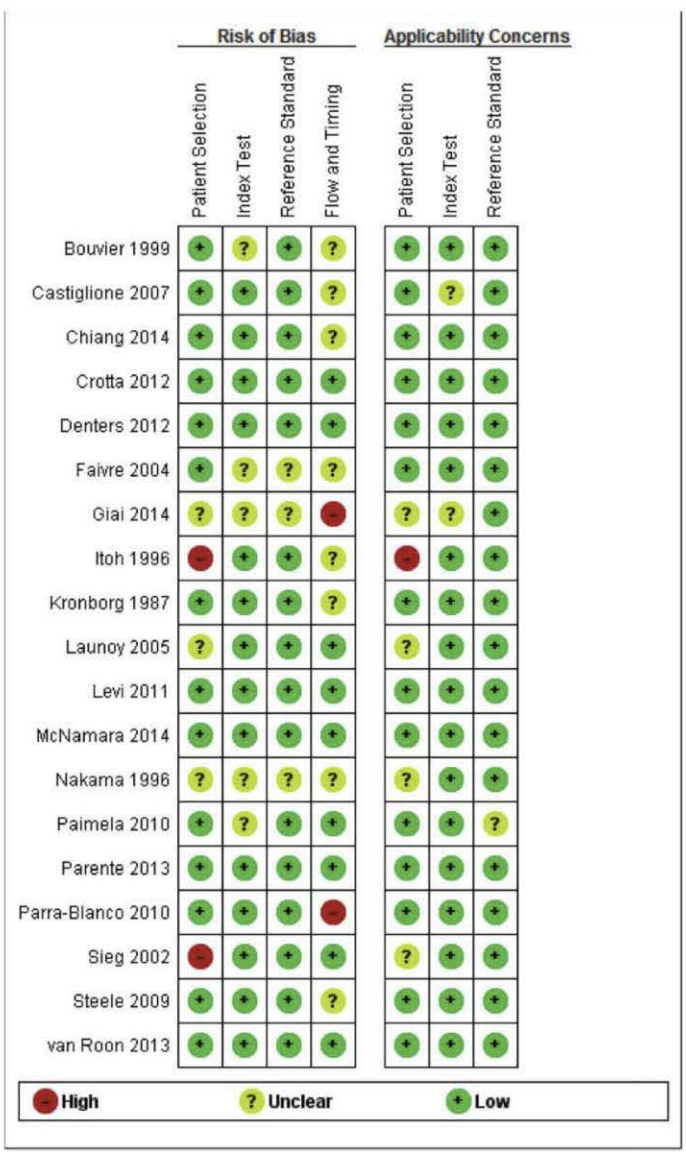


Figure 3b. *Type II studies. Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study*

studies had at least 1 year of follow up, five studies had a follow-up of three years, and one study had a follow-up of four years.

Twenty-one type I studies reported on advanced neoplasia (AN) as outcome; their median sample size was 1,046 (range 126 to 18,296). Figure 4 shows the Forrest plot of all included studies reporting on advanced neoplasia.

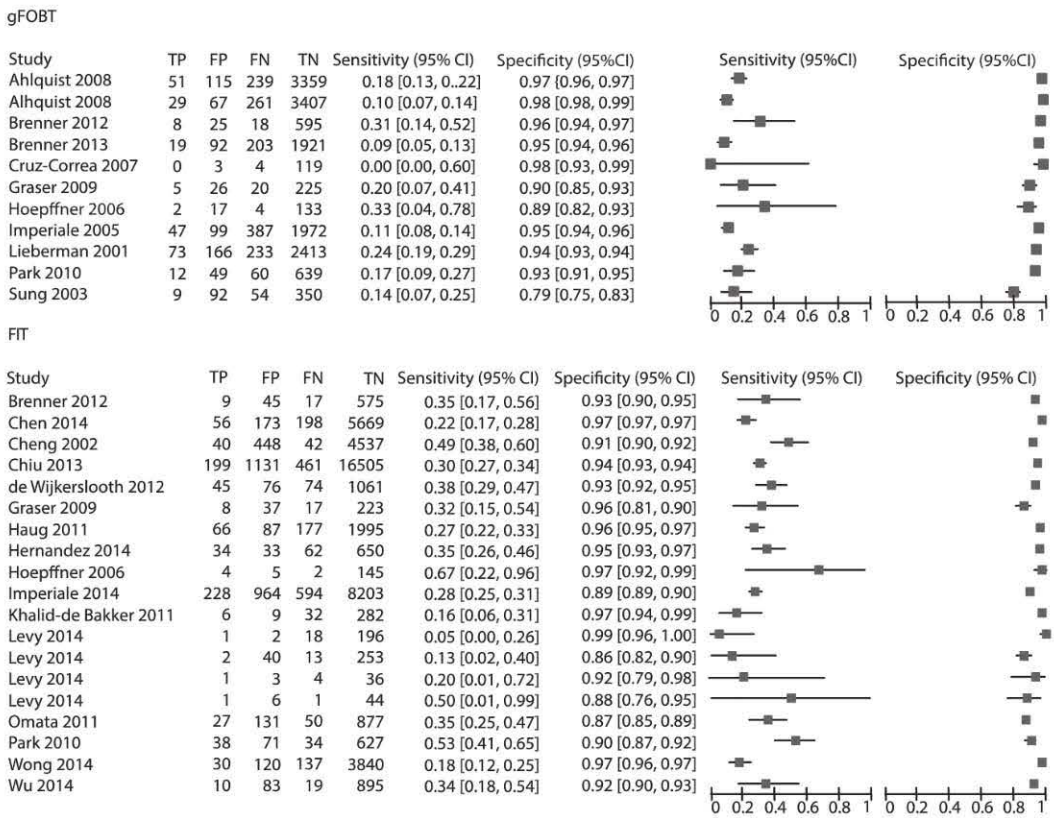


Figure 4. *Forrest plot of all gFOBT and FIT (Type I) for advanced neoplasia. For all FIT's a cut-off of 10 mcg Hb/ g faeces was used, unless this cut-off was unavailable.*

Sensitivities for detecting AN ranged from 0% to 33% for gFOBT and from 5% to 67% for FIT screening. Specificities ranged from 79% to 98% and from 86% to 99%, respectively. The cut-off for positivity of FIT varied between 2.4 to 50 µg Hb/g feces. The summary curve estimated by the HSROC model for all Type I studies for AN can be found in Figure 5. FIT showed a higher discriminative ability for AN than gFOBT (p=0.002).

In addition, sensitivities and specificities were calculated solely for those studies reporting on FIT screening with a cut-off value of 10 µg Hb/g feces and 20 µg Hb/g. Analyses for cut-off 10 µg Hb/g feces contained both qualitative as quantitative FITs. The sensitivity of FIT screening for detection of

AN ranged between 5% and 67% with a cut-off of 10 µg Hb/g, and from 13 to 44% with a cut-off of 20 µg Hb/g. The sensitivity for AN was lower for gFOBT screening with a summary sensitivity of 16% (95% CI 12-21%) compared to 31% 995% CI 25-39% for FIT with a cut-off of 10 µg Hb/g and 27% (95% CI 21-34%) with a cut-off of 20 µg Hb/g Specificities of FIT screening for detecting AN ranged from 87% to 97% for a cut-off of 10 µg Hb/g, and from 89% to 100% for a cut-off of 20 µg Hb/g. No significant differences in summary specificity for AN were found between gFOBT (94%; 95% CI 92-96%), FIT with a cut-off of 10 µg Hb/g (95%; 95% CI 92-97%) and with a cut-off of 20 µg Hb/g (97%; 95% CI 94-98%).

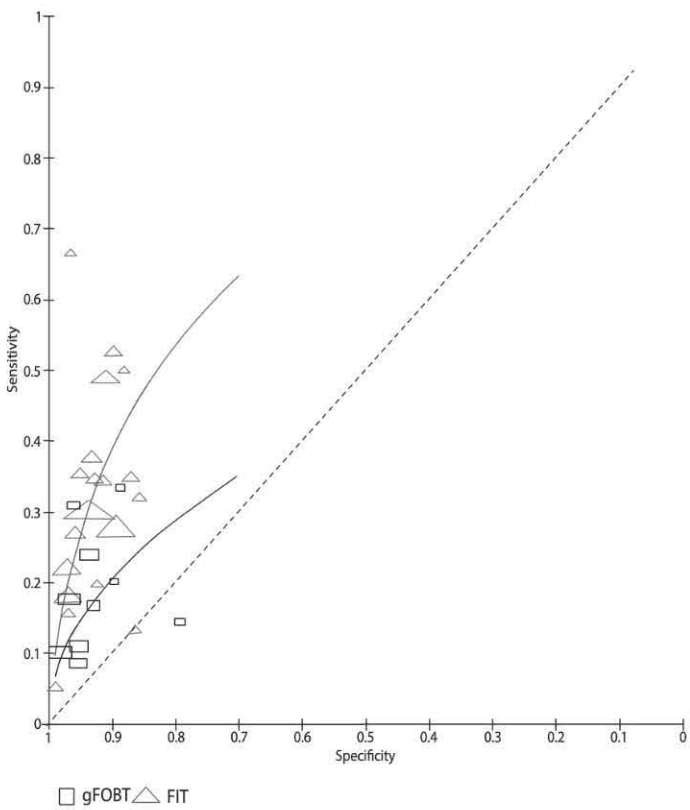


Figure 5. *Summary curve using the HSROC model for gFOBT and FIT (Type I) adjusting for multiple cut-offs for advanced neoplasia. Scale of individual study points is based on sample size.*

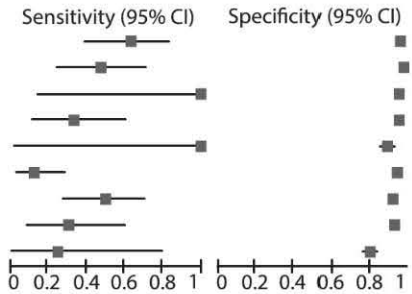
Type I studies - diagnostic test accuracy of gFOBT and FIT for colorectal cancer

Nineteen type I studies reported on colorectal cancer (CRC) as separate outcome measure; their median sample size was 2,235 (range 285 to 18,269). Figure 6 shows the Forrest plot of all included studies reporting on colorectal cancer. Sensitivities ranged from 13% to 100% for gFOBT, and from 0% to 100% for FIT. Specificities ranged from 80% to 98% for gFOBT and from 85% to 96% for FIT. The cut-off for positivity of FIT varied between 2.4 to 50 µg Hb/g feces. The summary curve estimated by the HSROC model for CRC can be found in Figure 7. FIT showed a higher discriminative ability for

CRC than gFOBT (p=0.025).

gFOBT

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Ahlquist 2008	12	154	7	3591	0.63 [0.38, 0.84]	0.96 [0.95, 0.97]
Ahlquist 2008	9	87	10	3658	0.47 [0.24, 0.71]	0.98 [0.97, 0.98]
Brenner 2012	2	31	0	613	1.00 [0.16, 1.00]	0.95 [0.93, 0.97]
Brenner 2013	5	106	10	2114	0.33 [0.12, 0.62]	0.95 [0.94, 0.96]
Graser 2009	1	30	0	245	1.00 [0.03, 1.00]	0.89 [0.85, 0.93]
Imperiale 2004	4	142	27	2332	0.13 [0.04, 0.30]	0.94 [0.93, 0.95]
Liebermann 2001	12	227	12	2634	0.50 [0.29, 0.71]	0.92 [0.91, 0.93]
Park 2010	4	57	9	690	0.31 [0.09, 0.61]	0.92 [0.90, 0.94]
Sung 2003	1	100	3	401	0.25 [0.01 0.81]	0.80 [0.76, 0.83]



FIT

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Brenner 2012	2	52	0	592	1.00 [0.16, 1.00]	0.92 [0.90, 0.94]
Chen 2014	9	220	4	5863	0.69 [0.39, 0.91]	0.96 [0.96, 0.97]
Cheng 2002	14	474	2	4577	0.88 [0.62, 0.98]	0.91 [0.90, 0.91]
Chiu 2013	22	1308	6	16960	0.79 [0.59, 0.92]	0.93 [0.92, 0.93]
de Wijkerslooth 2012	7	114	1	1134	0.88 [0.47, 1.00]	0.91 [0.89, 0.92]
Graser 2009	1	44	0	240	1.00 [0.03, 1.00]	0.85 [0.80, 0.89]
Haug 2011	11	142	3	2169	0.79 [0.49, 0.95]	0.94 [0.93, 0.95]
Hernandez 2014	5	62	0	712	1.00 [0.48, 1.00]	0.92 [0.90, 0.94]
Imperiale 2014	48	1144	17	8780	0.74 [0.61, 0.84]	0.88 [0.88, 0.89]
Levy 2014	0	7	0	45	Not estimable	0.87 [0.74, 0.94]
Levy 2014	0	3	1	213	0.00 [0.00, 0.97]	0.99 [0.96, 1.00]
Levy 2014	0	42	1	265	0.00 [0.00, 0.97]	0.86 [0.82, 0.90]
Levy 2014	0	4	0	40	Not estimable	0.91 [0.78, 0.97]
Nakama 2000	79	753	17	16815	0.82 [0.73, 0.89]	0.96 [0.95, 0.96]
Omata 2011	6	151	2	926	0.75 [0.35, 0.97]	0.86 [0.84, 0.88]
Park 2010	12	97	1	660	0.92 [0.64, 1.00]	0.87 [0.85, 0.89]
Wong 2014	2	148	10	3967	0.17 [0.02, 0.48]	0.96 [0.96, 0.97]
Wu 2014	3	90	2	912	0.60 [0.15, 0.95]	0.91 [0.89, 0.93]

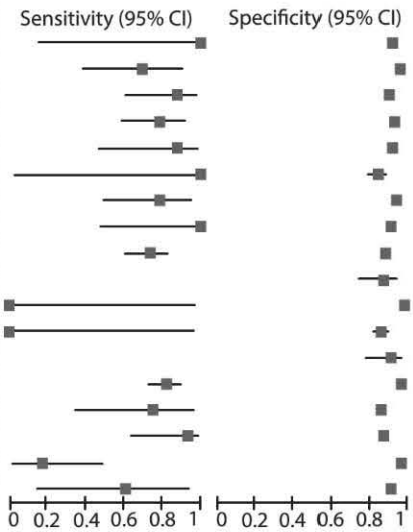


Figure 6. *Forrest plot of all gFOBT and FIT (Type I) for colorectal cancer. For all FIT's a cut-off of 10 µg Hb/ g faeces was used, unless this cut-off was unavailable.*

In addition, sensitivities and specificities were calculated solely for those studies reporting on a cut-off of 10 µg Hb/g feces and 20 µg Hb/g for FIT. Sensitivities for CRC ranged from 13% to 100% for gFOBT, from 0% to 100% for a FIT cut-off of 10 µg Hb/g, and from 50% to 100% for a cut-off of 20 µg Hb/g. Sensitivity for CRC was lower for gFOBT with a summary sensitivity of 41% (95% CI 29-54%), compared to 81% (95% CI 71-89%) for a FIT cut-off of 10 µg Hb/g, and 76% (95% CI 61-86%) for a FIT cut-off of 20 µg Hb/g. Specificities for FIT ranged from 87% to 99% when using a cut-off of 10 µg Hb/g, and from 88% to 96% with a cut-off of 20 µg Hb/g. No significant differences in summary specificity for colorectal cancer were found between gFOBT (94%; 95% CI 91-95%), FIT with a cut-off of 10 µg Hb/g (93%; 95% CI 91-95%), and with a cut-off of 20 µg Hb/g (93%; 95% CI 90-95%).

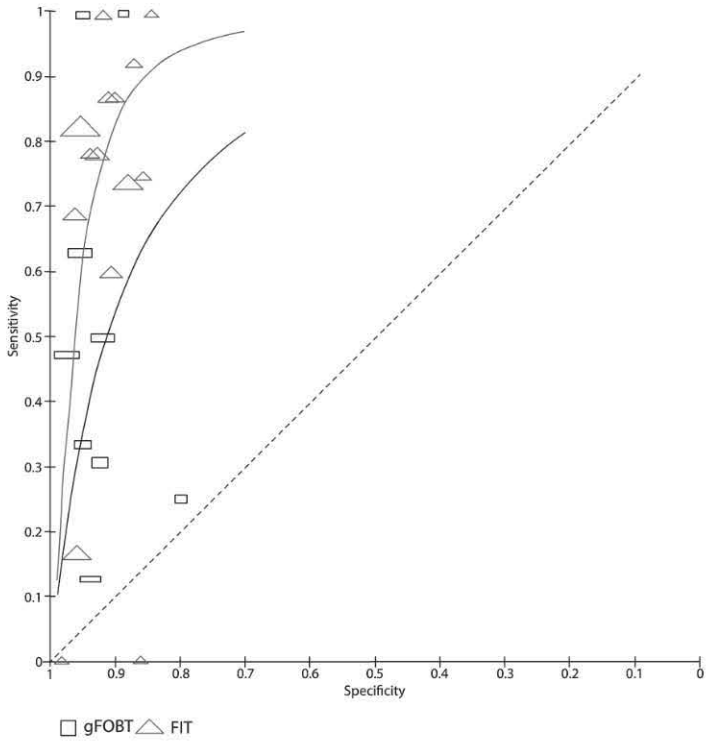


Figure 7. *Summary curve using the HSROC model for gFOBT and FIT (Type I) adjusting for multiple cut-offs for colorectal cancer. Scale of individual study points is based on sample size.*

Type I studies – linked ROC

Four studies (Brenner 2010, Brenner 2013 & Haug 2011 combined, Graser, 2009, Hoepffner 2006, Park 2010) compared FIT and gFOBT in the same population. The cut-off for positivity of FIT varied between 2.4 to 10 µg Hb/g feces. The summary curve estimated by the HSROC model for linked Type I studies for AN, can be found in Figure 8. FIT showed a higher discriminative ability for AN than gFOBT (p=0.073).

Type I studies heterogeneity analyses

There was a significant difference in sensitivity or specificity, or both, for males versus females for FIT, both for outcome AN as CRC (p<0.001). For gFOBT, difference in accuracy for males versus females was significant for outcome AN but not for CRC (p=0.002 and p=0.638, respectively). There was no evidence (all p-values >0.01) to suggest a difference in sensitivity or specificity, or both, between studies using one, two or three stools per screening round. There was no significant difference in sensitivity or specificity, or both, between studies using a quantitative or a qualitative FIT at a cut-off of 10 µg Hb/g for the outcome AN as well as CRC (p = 0.645 and p = 0.216, respectively).

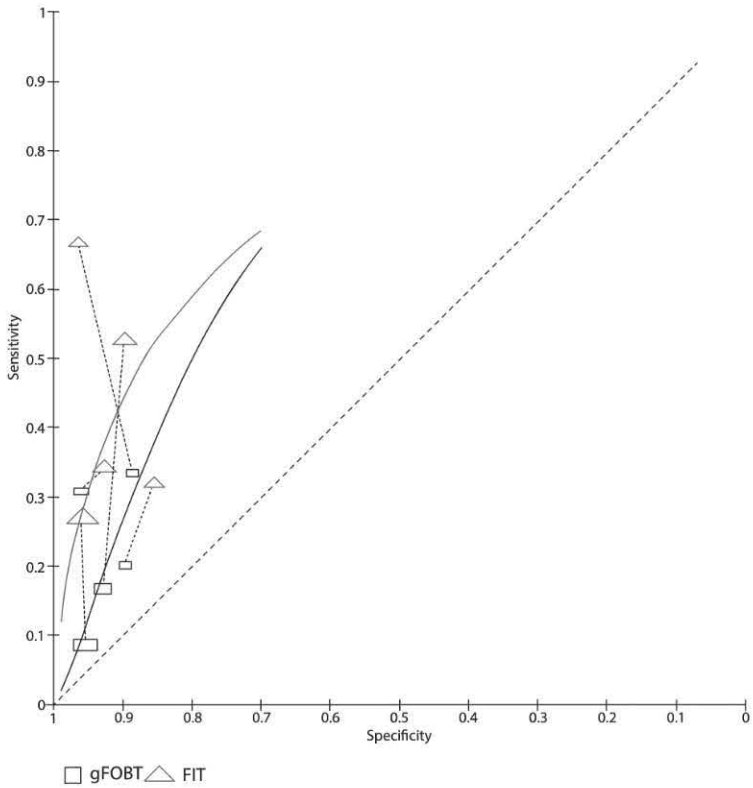


Figure 8. *Linked-HSROC curve of studies (Type I) with outcome advanced neoplasia (including: Brenner 2012, Graser 2009, Brenner 2013 and Haug 2011 combined, Park 2010, Hoepffner 2006). Scale of individual study points is based on sample size*

Type I studies sensitivity analyses

For the analyses including all cut-offs, a sensitivity analysis was undertaken by excluding the studies that yielded an high risk of bias following the QUADAS assessment (Alhquist 2008, Brenner 2013, Cruz-Correa 2007, Hoepffner 2006, Omata 2011, Sung 2013). Even when excluding these studies, FIT remained significantly superior to gFOBT in the HSROC model.

Type II studies - diagnostic test accuracy of gFOBT and FIT for colorectal cancer

There were nineteen type II studies that reported on CRC as separate outcome; their median sample size was 7,355 (range 1,179 to 747,076). Figure 9 shows the Forrest plot of all included studies reporting on CRC.

Sensitivities for detection of CRC ranged from 57% to 67% for gFOBT, and from 63% to 100% for FIT. Specificities ranged from 98% to 99% for gFOBT and from 91% to 97% for FIT. The cut-off for positivity of FIT varied between 5 to 250 µg Hb/g feces. The summary curve estimated by the HSROC model, including confidence ellipses for all Type I studies for AN, can be found in Figure 10. FIT showed a

higher discriminative ability for AN than gFOBT (p<0.001). In addition, sensitivities and specificities were calculated separately for those studies reporting on a FIT cut-off of 10 µg Hb/g feces, and those with a cut-off of 20 µg Hb/g for CRC. Sensitivities for detection of CRC ranged from 75% to 100% when using a cut-off of 10 µg Hb/g , and from 63% to 94% with a cut-off of 20 µg Hb/g. Sensitivity for CRC was lower for gFOBT with a summary sensitivity of 63% (95% CI 58-67%), compared to 87% (95% CI 80-92%) for a cut-off of 10 µg Hb/g, and 88% (95% CI 74-94%) for a cut-off of 20 µg Hb/g. Specificities for FIT ranged between studies from 92% to 96% when using a cut-off of 10 µg Hb/g, and from 92% to 97% for a cut-off of 20 µg Hb/g. Specificity for CRC was higher for gFOBT with a summary specificity for colorectal cancer of 98% (95% CI 98-99%) for gFOBT, compared to 92% (95% CI 92-95%) for FIT with a cut-off of 10 µg Hb/g, and 95% (95% CI 93-97%) at a cut-off of 20 µg Hb/g.

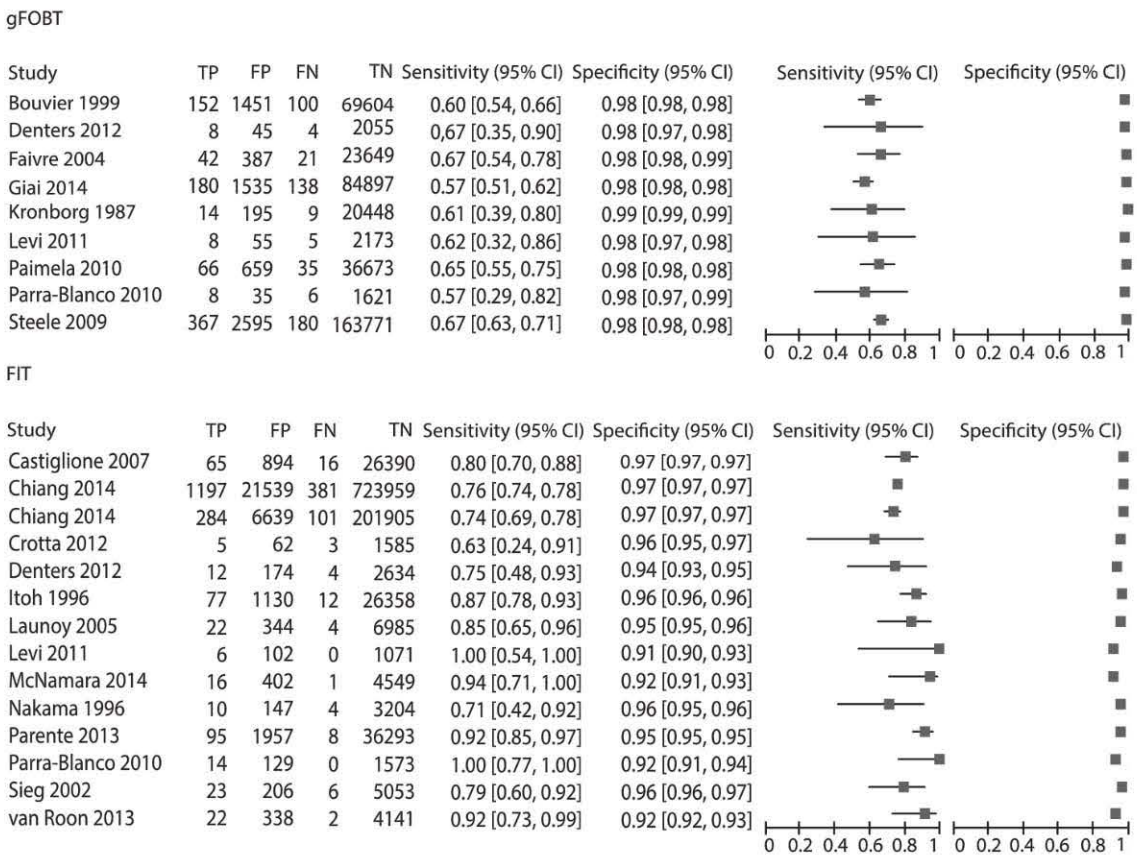


Figure 9. *Forrest plot of all gFOBT and FIT (Type II) for colorectal cancer. For all FIT's a cut-off of 10 mcg Hb/g faeces was used, unless this cut-off was unavailable.*

Type II studies heterogeneity analyses

There was a significant difference in sensitivity or specificity, or both, for males versus females for gFOBT, for the outcome CRC (p<0.001). There was no significant difference in sensitivity or specificity, or both, between studies using a quantitative or a qualitative FIT at a cut-off of 10 µg Hb/g for the

outcome CRC (p = 0.684). Heterogeneity related to the number of stools per screening round was not performed for gFOBT since all studies used three stools. For the following covariates analyses were not possible due to convergence difficulties; gender for FIT, number of stools for FIT.

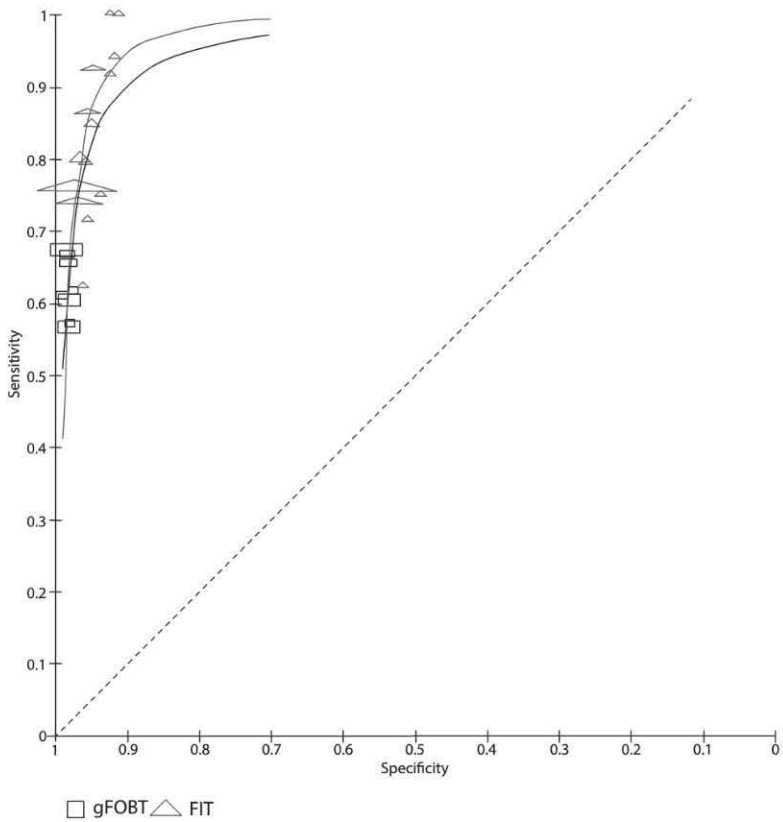


Figure 10. *Summary curve using the HSROC model for gFOBT and FIT (Type II) adjusting for multiple cut-offs for colorectal cancer. Scale of individual study points is based on sample size.*

Type II studies sensitivity analyses

For the analyses including all cut-offs, a sensitivity analysis was undertaken by excluding the studies that yielded an high risk of bias following the QUADAS assessment (Giai 2014, Itoh 1996 Nakama 1996, Parra-Blanco 2010, Sieg 2002)[71,72,76,79,82].Even when excluding these studies, FIT remained significantly superior to gFOBT in the HSROC model. The effect of removing studies, in which the percentage of participants with a positive FOBT that underwent the reference standard was unknown, was herein evaluated (Nakama 1996, Giai 2014)[71,76].

Table 3. Summary of findings.

Diagnostic accuracy of gFOBT compared to FIT					
Patients/ population	Asymptomatic, average-risk individuals over the age of 40 years undergoing colorectal cancer screening				
Prior testing	Only the results of the first screening round were included in this analysis				
Settings	Population- based colorectal cancer screening				
Index test	Guaiac faecal occult blood test or faecal immunochemical test				
Importance	Many screening programmes worldwide are currently changing from gFOBT to FIT-based screening				
Reference standard	Colonoscopy is the gold standard for the diagnosis of colorectal cancer which was used as the reference standard. Only in case a colonoscopy was not complete a CT-colonography (or double contrast barium enema) was used as a surrogate.				
Studies	Prospective and retrospective studies including average-risk individuals invited for colorectal cancer screening Type I: all screenees underwent both the index test and colonoscopy (n=23). Type II: only screenees with a positive index test underwent colonoscopy and all screen negatives were followed for at least one year (n=19).				
Quality concerns	Due to strict inclusion criteria most studies were of high quality. Few studies had unclear risk of bias due to poor reporting of a pre-specified cut-off value. Only three studies had a high risk of bias regarding the selection of study population. Regarding these studies, sensitivity analyses showed significant differences in outcome when excluding these studies from analyses.				
Test /subgroup*	studies (participants) gFOBT/FIT	summary sensitivity gFOBT (%, 95% CI)	summary specificity gFOBT (%, 95% CI	summary sensitivity FIT* (%, 95% CI)	summary specificity FIT* (%, 95% CI)
Type I					
advanced neoplasia	10 (15.741) / 16 (55.881)	15 (12-19)	94 (91-96)	31 (25-39)	95 (92-97)
colorectal cancer	8 (15.465) / 15 (69.998)	41 (29-54)	94 (91-96)	82 (71-89)	93 (90-95)
Type II					
colorectal cancer	9 (413.191) / 14 (1.082.153)	63 (58-67)	98 (98-99)	87 (80-92)	92 (92-95)
Conclusions	The results of this systematic review concludes that FIT is the preferred tool for FOBT-based population CRC screening due to the higher sensitivity and comparable similar specificity as compared to gFOBT.				

CAUTION: The results on this table should not be interpreted in isolation from the results of the individual included studies contributing to each summary test accuracy measure. These are reported in the main body of the text of the review.

* results for FIT cut-off 10 µg Hb/g faeces are shown.

Discussion

The main results are presented in the summary of findings (Table 3). For this review we chose to include two types of studies and report the results separately as they differ in type of study and yield comparable but not similar results. We included twenty-three type I studies and nineteen type II studies. Four type I studies evaluated the diagnostic accuracy of gFOBT, thirteen studies evaluated FIT and four studies assessed both gFOBT and FIT tests. Twenty-one studies evaluated the diagnostic accuracy for advanced neoplasia and nineteen studies evaluated the diagnostic accuracy for colorectal cancer. FIT showed a higher discriminative ability than gFOBT as assessed by the HSROC curve for both advanced neoplasia (p=0.002) and CRC (p=0.025). As type I studies allowed the use of multiple FIT cut-offs within one study population, the two most commonly used cut-offs worldwide (10 and 20 µg Hb/g feces) were analysed separately and these results were in line with the overall analyses. Nineteen type II studies were included, with six studies evaluating the diagnostic accuracy of gFOBT, ten studies evaluating FIT, and three studies combining both. In type II studies FIT also showed a higher discriminative ability than gFOBT as assessed by the HSROC curve for CRC (p<0.001). The results of this systematic review demonstrate that FIT is the preferred tool for FOBT-based population screening due to the superior sensitivity and similar specificity compared as to gFOBT screening. Furthermore, beside some qualitative tests, many FITs are quantitative tests which allows use of different cut-offs to tailor for screening resources and colonoscopy capacity. Finally, various studies have consistently shown that FIT screening is associated with higher uptake than gFOBT screening which is an important finding to reach a high coverage of the target population (i.e. cumulative uptake).

Strengths

The results of this review are based on strict and thorough searching without any language or date restrictions. The use of diagnostic test accuracy, or randomised controlled trial filters may lead to the loss of some studies, for this reason we have not used any filters[84]. Two independent reviewers identified and extracted data from the studies, thus decreasing inaccuracies related to single-person data extraction[85]. All included studies reported the results for average-risk, asymptomatic individuals after the age of 40 years, making our results reflective of a screening population. Also, data for different cut-offs were retrieved, in cases where this had not already been reported in the original publication, contacting the author provided additional data for most studies allowing sub-analyses with the two most commonly used cut-offs. These cut-offs were converted to the internationally used measuring standard of µg Hb/g feces[34]. To avoid potential bias caused by the use of an

inappropriate reference standard, e.g. barium-enema or sigmoidoscopy, we restricted the studies to those with colonoscopy as the reference standard. As mentioned above, two types of studies were included for this review and analysed separately. The inclusion of type I studies allowed evaluating both advanced neoplasia and colorectal cancer as outcome of FOBT-based screening. Advanced neoplasia is of special importance because by removing adenomas development of colorectal cancer and CRC deaths might be prevented[29,86]. In type II studies sensitivities and specificities were calculated with the use of interval carcinomas identified through adequate follow-up as a surrogate for the gold standard; colonoscopy. This different character of type I and type II studies may explain the observed differences in sensitivity and specificity. The use of interval carcinomas as endpoint in type II studies may underestimate the true proportion of false negatives, as by definition only those cancers were reported that had become clinically evident during the observation period. On the other hand, Type II studies are more reflective of a FOBT based screening programme in a general population. In these studies willingness to undergo FOBT as a primary screening tool was assessed whereas in type I studies participants had to be willing to undergo a full colonoscopy irrespective of the FOBT-result. For this reason type II studies are also often performed in larger populations. Combining both types of studies provides insight on both settings, and results in a broad evaluation of FOBT diagnostic test accuracy in colorectal cancer screening. Type I studies give insight in test sensitivity, whereas type II studies give insight in program sensitivity. The overall quality of included studies was high, supporting the validity of the results of our analyses.

Weaknesses

This systematic review was designed to evaluate the diagnostic test accuracy of two types of FOBTs commonly used for colorectal cancer screening. Even though diagnostic test accuracy is of major importance in screening, usability of the test and participation, e.g. willingness to undergo the screening test, is also very relevant. One major limitation of this review is that these latter points have not been taken into account for this review as they do not involve diagnostic test accuracy. Yet, these factors are also of importance when estimating screening efficacy on population level (programme sensitivity and specificity). The ultimate purpose of screening programmes is a decrease in colorectal cancer-related mortality. However, diagnostic test accuracy can only be used as a surrogate in estimating mortality decrease after screening. In past years, results of large prospective gFOBT-based screening trials have been published that unfortunately could not be included in this review as their main outcome was mortality. Mortality rates could not be converted into contingency tables to calculate sensitivity and specificity. We excluded non-interpretable test results and FOBT-positives who refused to undergo the reference standard from the 2x2 table and

in consequence from the meta-analysis. Missing FOBT results are likely to be completely random (incidental missing data) and will not lead to biased estimates of test accuracy. Because only the participants who received the reference standard were included in the type I studies analysis (complete case analysis) and positive participants who did not receive the reference standard in the type II studies were excluded from analysis, estimates of the accuracy of the diagnostic tests could be biased[87]. Because of the large size of the studies, we believe that excluding these participants for whom the reference standard result is missing is mostly at random and will not bias the results but will only decrease precision. We attempted to conduct a comprehensive search for studies, but the fact that the studies awaiting classification have not yet been incorporated may be a source of potential bias. Another limitation is the inability to explore the sources of heterogeneity concerning age, gender, ethnicity, adenoma type, tumour localization and stage distribution, because of limited information in the included studies. This problem was more prominent in type II studies, which often described large populations without prospective registration of study outcomes, but rather national databases of hospitals or cancer registries. The chosen random effect model accounts to a certain level for heterogeneity across studies by considering both within and between-study variation. This leads to a greater precision of the pooled estimates, but larger confidence[88].For only for a limited amount of studies was a direct comparison possible, as most studies did not perform both tests in the same patients. This could be a limitation, as results from non-comparative studies may differ from comparative studies[89].However, in this review results from the comparative studies are in line with the overall results. Many test brands are available, and sub analyses of these brands was not possible due to limited data and the use of different sub-types of the same brand.

A previously published meta-analysis evaluated the test accuracy of FIT for colorectal cancer but not advanced neoplasia and did not compare these results to the performance of gFOBT[90].To the best of our knowledge there is only one other systematic review comparing the diagnostic test accuracy of gFOBT and FIT[74].The highest summary sensitivity of this study for FIT (OC-sensor) was 87% and specificity was 93%, and for gFOBT (Hemoccult,) the summary sensitivity was 47% and summary specificity was 95%. They did not distinguish between type I, type II, and case-control studies, possibly leading to underestimation (type II studies) or overestimation (case-control) studies of test accuracy[30]. Even though this large variation in type of studies included, their search strategy was very limited yielding only 761 hits and 22 inclusions. Finally, only three brands of FOBTs were included in their meta-analysis. In general this is the first review to adequately and systematically present an evaluation of gFOBT and FIT screening in an average-risk population.

Applicability of findings to the review question

All participants included in this review were asymptomatic, average-risk individuals over the age of 40 years old, and invited for colorectal cancer screening, making the findings of this review extremely relevant for colorectal cancer screening programmes. Two types of studies were included. Type I studies are more homogenous than type II studies, yet may be less representative of a FOBT-based screening population. This is due to the fact that all screenees had to be willing to undergo colonoscopy. For type II studies false negatives were identified through interval carcinomas that were identified during follow-up. This might give an underestimation of test sensitivity, yet is appropriately representative of FOBT population-based screening programmes.

Implications for practice

Fecal immunochemical testing has a superior sensitivity compared to guaiac fecal occult blood testing and is the preferred method of occult blood screening in terms of diagnostic test accuracy. Test usability and participation were not evaluated for the purpose of this review. The summary of findings table should be interpreted with acknowledgement of this. However, FIT's quantitative nature allows the use of different cut-offs tailoring to screening resources and colonoscopy capacity[24]. Furthermore, it should be noted that several studies consistently reported higher rates of participation for FIT than for gFOBT screening[91,92]. Both gFOBT and FIT have lower sensitivity for colorectal cancer than colonoscopy as gold standard. However, when combining test accuracy with participation FIT-based screening in many populations results in a higher diagnostic yield of advanced neoplasia compared to other CRC screening methods[10,28].

Implications for research

Future studies should be conducted in a prospective manner mimicking population-based colorectal cancer screening and targeting average-risk populations. We encourage authors to systematically report data on participation, positivity rate and colonoscopy adherence. Also, future studies should report a clear definition of advanced neoplasia and interval carcinomas. In the included studies definitions of interval carcinomas were often vague or completely missing. The ultimate purpose of colorectal cancer screening is decreasing mortality, so future studies should be conducted to compare long-term follow up on mortality between gFOBT- and FIT-based colorectal cancer screening programmes.

References

1. Andermann A, Blancquert I, Beauchamp S, Dery V. Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. Bull World Health Organ 2008; 86: 317-9.
2. Wilson JM, Jungner YG. [Principles and practice of mass screening for disease}. Principios y metodos del examen colectivo para identificar enfermedades. Bol Oficina Sanit Panam 1968; 65(4): 281-393.
3. Ciatto S, Martinelli F, Castiglione G, et al. Association of FOBT-assessed fecal Hb content with colonic lesions detected in the Florence screening programme. Br J Cancer 2007; 96: 218-21.
4. Edwards JB. Screening for colorectal cancer using fecal blood testing: varying the positive cut-off value. Pathology 2005; 37: 565-8.
5. Levi Z, Rozen P, Hazazi R, et al. A quantitative immunochemical fecal occult blood test for colorectal neoplasia. Ann Intern Med 2007; 146: 244-55.
6. Rozen P, Waked A, Vilkin A, Levi Z, Niv Y. Evaluation of a desk top instrument for the automated development and immunochemical quantification of fecal occult blood. Med Sci Monit 2006; 12: MT27-MT32.
7. van Doorn SC, Stegeman I, Stroobants AK, et al. Fecal immunochemical testing results and characteristics of colonic lesions. Endoscopy 2015; 47(11): 1011-7.
8. van Rijn JC, Reitsma JB, Stoker J, Bossuyt PM, van Deventer SJ, Dekker E. Polyp miss rate determined by tandem colonoscopy: a systematic review. Am J Gastroenterol 2006; 101: 343-50.
9. Young GP. Population-based screening for colorectal cancer: Australian research and implementation. J Gastroenterol Hepatol 2009; 24 Suppl 3: S33-S42.
10. Kuipers, Ernst J, Grady, et al. Colorectal cancer. Nature Reviews Disease Primers 2015; 1: 15066.
11. Mandel JS, Bond JH, Church TR, et al. Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota Colon Cancer Control Studyh. N Engl J Med 1993; 328: 1365-71.
12. Lieberman DA, Weiss DG, Veterans Affairs Cooperative Study G. One-time screening for colorectal cancer with combined fecal occult-blood testing and examination of the distal colon. New England Journal of Medicine 2001; 345(8): 555-60.
13. Kronborg O, Fenger C, Olsen J, Jorgensen OD, Sondergaard O. Randomised study of screening for colorectal cancer with fecal-occult-blood test. Lancet 1996; 348: 1467-71.
14. Hardcastle JD, Chamberlain JO, Robinson MH, et al. Randomised controlled trial of fecal-occult-blood screening for colorectal cancer. Lancet 1996; 348: 1472-7.
15. Lindholm E, Brevinge H, Haglind E. Survival benefit in a randomized clinical trial of fecal occult blood screening for colorectal cancer. Br J Surg 2008; 95(8): 1029-36.
16. 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. Am J Gastroenterol 2008; 103(6): 1541-9.
17. Shaukat A, Mongin SJ, Geisser MS, et al. Long-term mortality after screening for colorectal cancer. N Engl J Med 2013; 369(13): 1106-14.
18. Allison JE, Tekawa IS, Ransom LJ, Adrain AL. A comparison of fecal occult-blood tests for colorectal-cancer screening. N Engl J Med 1996; 334: 155-9.
19. Young GP, St John DJ, Winawer SJ, Rozen P. Choice of fecal occult blood tests for colorectal cancer screening: recommendations based on performance characteristics in population studies: a WHO (World Health Organization) and OMED (World Organization for Digestive Endoscopy) report. Am J Gastroenterol 2002; 97: 2499-507.

20. Brenner H, Tao S, Haug U. Low-dose aspirin use and performance of immunohcmical fecal occult blood tests. JAMA 2010; 304(22): 2513-20.
21. Levi Z, Rozen P, Hazazi R, et al. Sensitivity, but not specificity, of a quantitative immunochemical fecal occult blood test for neoplasia is slightly increased by the use of low-dose aspirin, NSAIDs, and anticoagulants. Am J Gastroenterol 2009; 104: 933-8.
22. Chiang TH, Chuang SL, Chen SL, 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: 1317-26.
23. Hundt S, Haug U, Brenner H. Comparative evaluation of immunochemical fecal occult blood tests for colorectal adenoma detection. Ann Intern Med 2009; 150(3): 162-9.
24. Wilschut JA, Habbema JD, van Leerdam ME, et al. Fecal occult blood testing when colonoscopy capacity is limited. J Natl Cancer Inst 2011; 103(23): 1741-51.
25. Zorzi M, Fedeli U, Schievano E, et al. Impact on colorectal cancer mortality of screening programmes based on the fecal immunochemical test. Gut 2015; 64(5): 784-90.
26. Schreuders EH, Grobbee EJ, Spaander MCW, Kuipers EJ. Advances in fecal tests for colorectal cancer screening. Curr Treat Options Gastroenterol 2016; 14(1): 152-62.
27. Spada C, Barbaro F, Andrisani G, et al. Colon capsule endoscopy: What we know and what we would like to know. World J Gastroenterol 2014; 20(45): 16948-55.
28. Schreuders EH, Ruco A, Rabeneck L, et al. Colorectal cancer screening: a global overview of existing programmes. Gut 2015; 64(10): 1637-49.
29. Winawer SJ, Fletcher RH, Miller L, et al. Colorectal cancer screening: clinical guidelines and rationale. Gastroenterology 1997; 112: 594-642.
30. Deeks JJ, Bossuyt PM, Gatsonis C. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy, Version 1.0.0: The Cochrane Collaboration; 2009.
31. Leeflang MM, Bossuyt PM, Irwig L. Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. J Clin Epidemiol 2009; 62(1): 5-12.
32. Lansdorp-Vogelaar I, von Karsa L, International Agency for Research on Cancer. European guidelines for quality assurance in colorectal cancer screening and diagnosis. First Edition--Introduction. 2012; 44(Suppl 3): SE15-SE30.
33. Sanduleanu S, le Clercq CM, Dekker E, et al. Definition and taxonomy of interval colorectal cancers: a proposal for standardising nomenclature. Gut 2015; 64(8): 1257-67.
34. Fraser CG, Allison JE, Halloran SP, Young GP, Expert Working Group on Fecal Immunochemical Tests for Hemoglobin CCSCWEO. A proposal to standardize reporting units for fecal immunochemical tests for hemoglobin. J Natl Cancer Inst 2012; 104(11): 810-4.
35. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011; 155(8): 529-36.
36. Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter10: Analysing and Presenting Results. In: Deeks Jj BPMGC, ed. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1 0: The Cochrane Collaboration. Available from: http://srdta.cochrane.org/; 2010.
37. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med 2001; 20: 2865-84--84.
38. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-

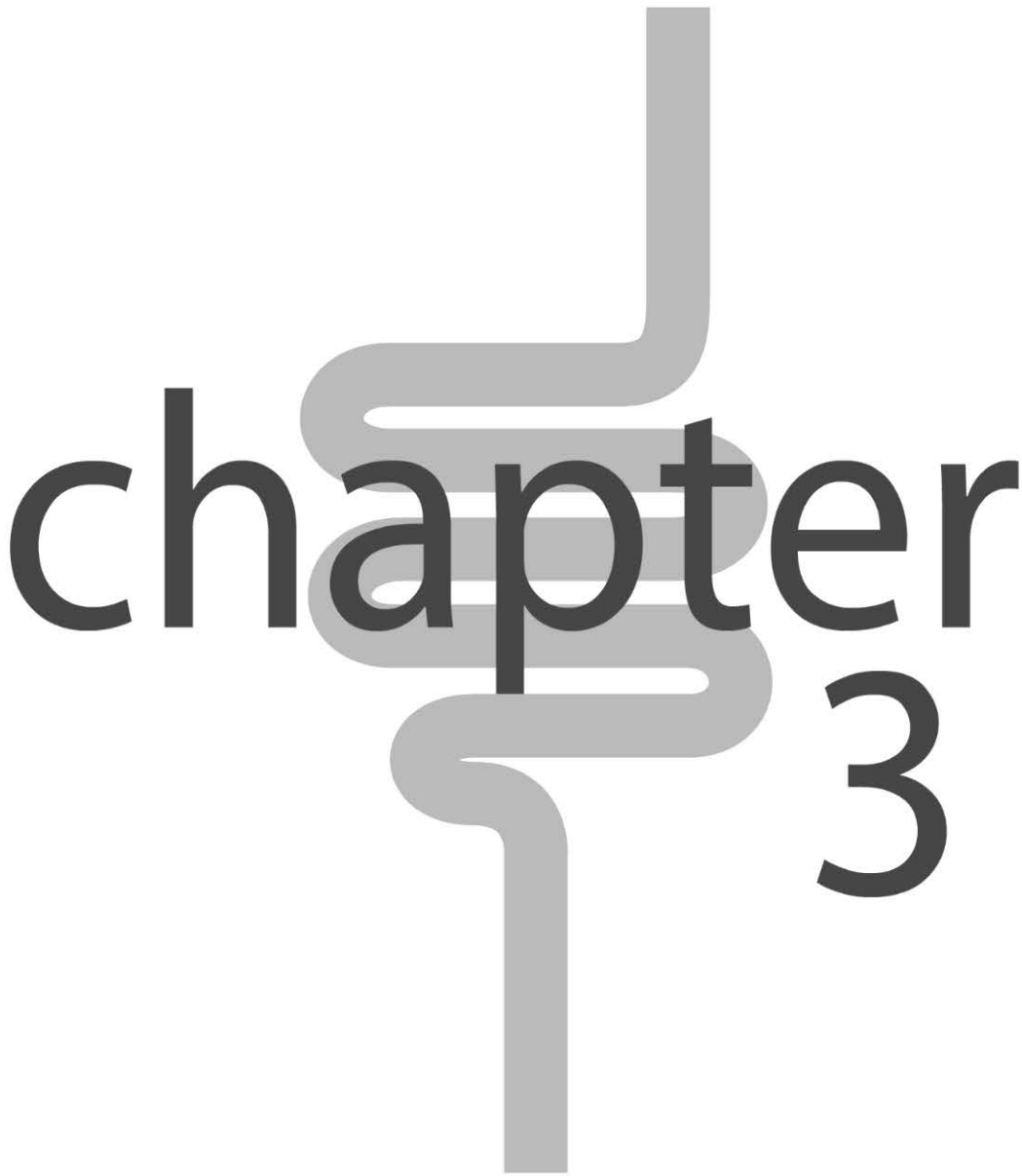
regression. Stat Med 2002; 21(4): 589-624.

39. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis ofsensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 2005; 58: 982-90--90.
40. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003; 327(7414): 557-60.
41. Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM. Systematic reviews of diagnostic test accuracy. Ann Intern Med 2008; 149: 889-97.
42. Song F, Khan KS, Dinnes J, Sutton AJ. Asymmetric funnel plots and publication bias in meta-analyses of diagnostic accuracy. Int J Epidemiol 2002; 31: 88-95.
43. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol 2005; 58: 882-93.
44. Faivre, Dancourt, e Lejeun, et al. Reduction in Colorectal Cancer Mortality by Fecal occult Blood Screening in a French controlled study. Gastroenterology 2004; 126(7): 1674-80.
45. European Colorectal Cancer Screening Guidelines Working G, von Karsa L, Patnick J, 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.
46. Haug U, Kuntz KM, Knudsen AB, Hundt S, Brenner H. Sensitivity of immunochemical fecal occult blood testing for detecting left-vs right-sided colorectal neoplasia. Br J Cancer 2011; 104(11): 1779-85.
47. Ahlquist DA, Sargent DJ, Loprinzi CL, et al. Stool DNA and occult blood testing for screen detection of colorectal neoplasia. Annals of Internal Medicines 2008; 149(7): 441-50.
48. Brenner G, Faure H, Heuer S, Reinholz J. Detection of colorectal findings for cancer prevention by immunochemical stool test with different sensitivity levels. Zeitschrift fur gastroenterologie 2012; 50(10): 1083-8.
49. Chen, Y Y, Chen TH, et al. Accuracy of immunochemical fecal occult blood test for detecting colorectal neoplasms in individuals undergoing health check-ups. Advances in Digestive Medicine 2014; 1: 74-9.
50. Cheng, Wong, Hong, et al. Colorectal cancer screening in asymptomatic adults: Comparison of colonoscopy, sigmoidoscopy and fecal occult blood tests. J Formos Med Assoc 2002; 101(10): 685-90.
51. Chiu, Lee, Tu, et al. Association between early stage colon neoplasms and false-negative results from the fecal immunochemical test. Clin Gastroenterol Hepatol 2013; 11(7): 832-8.
52. Cruz-Correa M, Schultz K, Jagannath S, et al. Performance characteristics and comparison of two fecal occult blood tests in patients undergoing colonoscopy. Dig Dis Sci 2007; 52(4): 1009-13.
53. De Wijkerslooth, Stoop, t Bossuy, et al. Immunochemical fecal occult blood testing is equally sensitive for proximal and distal advanced Neoplasia. Am J Gastroenterol 2012; 107(10): 1570-8.
54. Graser, Stieber, Nagel, et al. Comparison of CT colonography, colonoscopy, sigmoidoscopy and fecal occult blood tests for the detection of advanced adenoma in an average risk population. Gut 2009; 58(2): 241-8.
55. Hernandez V, Cubiella J, Gonzalez-Mao MC, et al. Fecal immunochemical test accuracy in average-risk colorectal cancer screening. 2014.
56. Hoepffner N, Shastri YM, Hanisch E, et al. Comparative evaluation of a new bedside fecal occult blood test in a prospective multicentre study. Aliment Pharmacol Ther 2005; 23: 145-54.
57. Imperiale, Ransohoff, Itzkowitz, et al. Fecal DNA versus fecal occult blood for colorectal-cancer screening in an average-risk population. N Engl J Med 2004; 351(26): 2704-14.
58. Imperiale TF, Ransohoff DF, Itzkowitz SH, et al. Multitarget stool DNA testing for colorectal-cancer screening. New Engl J

Med 2014; 370: 1287-97.

59. Khalid-de Bakker CA, Jonkers DM, Sanduleanu S, et al. Test performance of immunologic fecal occult blood testing and sigmoidoscopy compared with primary colonscopy screening for colorectal advanced adenomas. Cancer Prev Res 2011; 4: 1563-71.
60. Levy BT, Bay C, Xu Y, et al. Test characteristics of fecal immunochemical tests (FIT) compared with optical colonoscopy. J Med Screen 2014; 21(3): 133-43.
61. Nakama, Zhang, Fattah, et al. Colorectal cancer in iron deficiency anemia with a positive result on immunochemical fecal occult blood. International Journal of Colorectal Disease 2000; 15: 271-4.
62. Omata, Shintani, Isozaki, et al. Diagnostic performance of quantitative fecal immunochemical test and multivariate prediction model for colorectal neoplasms in asymptomatic individuals. Eur J Gastroenterol Hepatol 2011; 23(11): 1036-41.
63. Park, Ryu, Kim, et al. Comparison of guaiac-based and quantitative immunochemical fecal occult blood testing in a population at average risk undergoing colorectal cancer screening. Am J Gastroenterol 2010; 105(9): 2017-25.
64. Sung, Chan, Leung, et al. Screening for colorectal cancer in Chinese: comparison of fecal occult blood test, flexible sigmoidoscopy, and colonoscopy. Gastroenterology 2003; 124(3): 608-14.
65. Wong MCS, Ching JYL, Chan VCW, et al. Should prior FIT results be incorporated as an additional variable to estimate risk of colorectal neoplasia? A prospective study of 5,813 screening colonoscopies. Plos One 2014; 9(12): e114332-e.
66. Wu T, Kuo K, Wu Y, K Lin. Diagnostic accuracy of a single qualitative immunochemical fecal occult blood test coupled with physical measurements. Chin Med J (Engl) 2014; 127(24): 4164-70.
67. Bouvier V, Launoy G, Herbert C, Lefevre H, Maurel J, Gignoux M. Colorectal cancer after a negative Hemeoccult II(R) test and programme sensitivity after a first round of screening: The experience of the Department of Calvados (France). British Journal of Cancer 1999; 81: 305-9.
68. Castiglione G, Visioli C B, Ciatto S, et al. Sensitivity of latex agglutination fecal occult blood test in the Florence District population-based colorectal cancer screening programme. Br J Cancer 2007; 96: 1750-4.
69. Crotta S, Segnan N, Paganin S, Dagnes B, Rosset R, Senore C. High rate of advanced adenoma detection in 4 rounds of colorectal cancer screening with the fecal immunochemical test. Clin Gastroenterol Hepatol 2012; 10(6): 633-8.
70. Denters M J, Deutekom M, Bossuyt P M, Stroobants A K, Fockens P, Dekker E. Lower risk of advanced neoplasia among patients with a previous negative result from a fecal test for colorectal cancer. Gastroenterology 2012; 142: 497-504.
71. Giai J, Exbrayat C, Boussat B, et al. Sensitivity of a colorectal cancer screening program based on a guaiac test: a population-based study. Clin Res Hepatol Gastroenterol 2014; 38(1): 106-11.
72. Itoh M, Takahashi K, Nishida H, Sakagami K, Okubo T. Estimation of the optimal cut off point in a new immunological fecal occult blood test in a corporate colorectal cancer screening programme. J Med Screen 1996; 3: 66-71.
73. Kronborg O, Fenger C, Sondergaard O, Pedersen KM, Olsen J. Initial mass screening for colorectal cancer with fecal occult blood test. A prospective randomized study at Funen in Denmark. Scand J Gastroenterol 1987; 22(6): 677-86.
74. Launoy GD, Bertrand HJ, C. B. Evaluation of an immunochemical fecal occult blood test with automated reading in screening for colorectal cancer in a general average-risk population. Int J Cancer 2005; 115(3): 493-6.
75. Levi Z, Birkenfeld S, Vilkin A, et al. A higher detection rate for colorectal cancer and advanced adenomatous polyp for screening with immunochemical fecal occult blood test than guaiac fecal occult blood test, despite lower compliance rate. A prospective, controlled, feasibility study. Int J Cancer 2011; 128: 2415-24.
76. Nakama H, Kamijo N, Abdul Fattah A S, Zhang B. Validity of immunological fecal occult blood screening for colorectal cancer: a follow up study. J Med Screen 1996; 3: 63-5.

77. McNamara D, Leen R, Seng-Lee C, et al. Sustained participation, colonoscopy uptake and adenoma detection rates over two rounds of the Tallaght-Trinity College colorectal cancer screening programme with the fecal immunological test. Eur J Gastroenterol Hepatol 2014; 26(12): 1415-21.
78. Paimela H, Malila N, Palva T, Hakulinen T, Vertio H, Jarvinen H. Early detection of colorectal cancer with fecal occult blood test screening. Br J Surg 2010; 97: 1567-71.
79. Parra-Blanco A, Gimeno-Garcia AZ, Quintero E, et al. Diagnostic accuracy of immunochemical versus guaiac fecal occult blood tests for colorectal cancer screening. J Gastroenterol 2010; 45(7): 703-12.
80. Parente F, Boemo C, Ardizzoia A, et al. Outcomes and cost evaluation of the first two rounds of a colorectal cancer screening program based on immunochemical fecal occult blood test in northern Italy. Endoscopy 2013; 45: 27-34.
81. Steele R J C, McClements P L, Libby G, et al. Results from the first three rounds of the Scottish demonstration pilot of FOBT screening for colorectal cancer. Gut 2009; 58: 530-5.
82. Sieg A, Wirth A, Luthgens K, Schmidt-Gayk H. Six years of screening for colorectal neoplasms with an immunological fecal occult hemoglobin test. Verdauungskrankheiten 2002; 20: 114-7.
83. van Roon AH, Goede SL, van Ballegooijen M, et al. Random comparison of repeated fecal immunochemical testing at different intervals for population-based colorectal cancer screening. Gut 2013; 62(3): 409-15.
84. Doust JA, Pietrzak E, Sanders S, Glasziou PP. Identifying studies for systematic reviews of diagnostic tests was difficult due to the poor sensitivity and precision of methodologic filters and the lack of information in the abstract. J Clin Epidemiol 2005; 58(5): 444-9.
85. Buscemi N, Hartling L, Vandermeer B, Tjosvold L, Klassen TP. Single data extraction generated more errors than double data extraction in systematic reviews. J Clin Epidemiol; 7(697-703).
86. Zauber AG, Winawer SJ, O'Brien MJ, et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. N Engl J Med 2012; 366(8): 687-96.
87. Naaktgeboren CA, de Groot JA, Rutjes AW, Bossuyt PM, Reitsma JB, Moons KG. Anticipating missing reference standard data when planning diagnostic accuracy studies. BMJ 2016; 352: i402-i.
88. Lijmer JG, Bossuyt PM, Heisterkamp SH. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. Stat Med 2002; 21(11): 1525-37.
89. Takwoingi Y, Leeflang MM, Deeks JJ. Empirical evidence of the importance of comparative studies of diagnostic test accuracy. 2013 2013; 7(544-554).
90. Lee JK, Liles EG, Bent S, Levin TR, Corley DA. Accuracy of fecal immunochemical tests for colorectal cancer: systematic review and meta-analysis. Ann Intern Med 2014; 160(3): 171.
91. Hol L, Wilschut JA, van Ballegooijen M, et al. Screening for colorectal cancer: random comparison of guaiac and immunochemical fecal occult blood testing at different cut-off levels. Br J Cancer 2009; 100(7): 1103-10.
92. van Rossum LG, van Rijn AF, Laheij RJ, et al. Random comparison of guaiac and immunochemical fecal occult blood tests for colorectal cancer in a screening population. Gastroenterology 2008; 135(1): 82-90.


chapter
3

Prospective comparison of three different colorectal cancer screening strategies: colonoscopy, flexible sigmoidoscopy and multiple rounds of FIT

Esmée J. Grobbee, Manon van der Vlugt*, Anneke J. van Vuuren, An K. Stroobants, Rosalie C. Mallant-Hent, Iris Lansdorp-Vogelaar, Patrick M. Bossuyt, Ernst J. Kuipers, Evelien Dekker, Manon C.W. Spaander*

**both authors contributed equally*

Submitted

Abstract

Background Several modalities for colorectal cancer (CRC) screening are available; these in particular include colonoscopy, flexible sigmoidoscopy (FS) and fecal immunochemical testing (FIT). No literature is available on the comparison of endoscopic screening and multiple rounds of FIT screening.

Methods We compared three CRC screening strategies involving 30,007 asymptomatic persons aged 50-74 years. Selected persons were either invited for four rounds of FIT (n=15,046), for once-only FS (n=8,407), or for once-only colonoscopy (n=6,600). Primary outcome was diagnostic yield (DY) of advanced neoplasia (AN), evaluated in an intention-to-screen analysis (defined as number of screenees with AN detected relative to all invitees) and in an as-screened analysis (defined as number of screenees with AN detected relative to all participants).

Results Participation rate was significantly higher for FIT screening (77%) than for FS (31%; p<0.001) and colonoscopy (24%; p<0.001). Number of colonoscopies performed relative to eligible invitees was higher for colonoscopy (24%) compared to FIT (13%; p<0.001) and FS (3%; p<0.001). In the intention-to-screen analysis, the DY for AN was higher with FIT-screening (4.5%; 95% CI 4.2-4.9), compared to colonoscopy (2.2%; 1.8-2.6) and FS (2.3%; 2.0-2.7). In the as-screened analysis, detection of AN was higher for endoscopic screening: 9.1% (7.7-10.7) with colonoscopy and 7.4% (6.5-8.5) with FS, compared to 6.1% with FIT (5.7-6.6). CRC detection among participants was similar for all three strategies. Number needed to screen to detect one AN was lower for FIT: 22 screenees, compared to 43 for FS and 46 for colonoscopy.

Conclusion Multiple-round FIT screening detects significantly more AN on a population level, compared to once-only FS and colonoscopy screening, and with significantly fewer colonoscopies needed.

Introduction

Colorectal cancer (CRC) is the third most common malignancy in the world. Its incidence increases, both as a result of expansion of the elderly population as well as an increase in risk factors[1]. The implementation of CRC screening programs has increased substantially over the past decades. A range of screening methods is available, varying in invasiveness and diagnostic accuracy. It is yet unclear which modality has the largest effect on CRC-related morbidity and mortality[2]. At present, screening programs differ markedly around the world, with colonoscopy, flexible sigmoidoscopy and faecal occult blood testing as the most commonly used screening strategies[3].
Colonoscopy as a first-line screening tool is more popular in the United States, whereas most countries in Europe and Asia prefer a two-step approach using less invasive screening methods, such as fecal occult blood testing or sigmoidoscopy as a first screen, followed by a colonoscopy in case of a positive test[3]. This two-step approach has the advantage of potentially higher participation rates and a lesser demand on colonoscopy capacity, but the disadvantage of lower sensitivity for detection of advanced neoplasia.

Annual or biennial screening with guaiac fecal occult blood testing has been shown to reduce CRC mortality rates[11,12]. At, present, the use of the guaiac test is largely replaced by FIT, as quantitative test for hemoglobin, FIT has a higher diagnostic accuracy for AN and is easier to use[13-16]. Nevertheless, FIT still has a relatively low sensitivity for CRC, and especially advanced adenomas, with reported cancer miss rates of 25% in a single round of FIT-screening[17]. Consequently, the effectiveness of a FIT-screening program is highly dependent on participation and adherence to repeat screening, e.g. longitudinal adherence[18].

As the majority of advanced neoplasia is located in the rectum and sigmoid colon, flexible sigmoidoscopy is considered a suitable method for CRC screening[8]. Several large randomized trials reported corresponding CRC-related mortality reductions ranging between 31% and 38% [8-10]. However, concerns have been raised about missing proximal advanced neoplasia, and the subsequent limited efficacy in reducing proximal CRC.

Colonoscopy is currently considered as the reference standard for diagnosing CRC and its precursor lesions, colorectal adenomas. It allows for simultaneous inspection of the colon and removal of lesions. Currently, no literature is available on the effect on mortality rates of primary colonoscopy screening in randomized trials. For that purpose, three randomized controlled colonoscopy screening

trials started a few years ago, of which the first results are expected after 2020[4-6]. A systematic review and meta-analysis of observational colonoscopy studies suggested that colonoscopy led to a significant mortality reduction with a summary estimate of 68% [7]. While colonoscopy is highly effective in detecting CRC and colorectal adenomas, it is also an invasive procedure associated with patient discomfort, complication rates and substantial costs.

Several studies have compared endoscopic screening strategies to FIT and find higher uptakes for FIT screening yet lower detection rates for advanced neoplasia[4-6].However, these studies comprise only one round of screening, while true the impact of FIT-screening is attained over multiple rounds. For this reason we aimed to compare the diagnostic yield of once-only colonoscopy, once-only FS and four rounds of FIT in population-based CRC screening, including interval cancer rate. As it is of key importance for policy makers to know the impact of different screening programs over multiple rounds with long-term follow up.

Methods

Study population and design

For the purpose of this study we combined the results of three population trials, each comprising randomly selected, screening-naive persons from the same source population. The design of these trials has been described previously [19-22].

In short, 30,052 asymptomatic persons aged 50 to 75 years living in the west of the Netherlands were invited for CRC screening. These persons were randomly selected from municipal registers, sorted according to household and were randomized before invitation. Those allocated to FIT-screening were invited between November 2006 and December 2014 for four rounds of FIT-screening with two-year intervals. Those allocated to flexible sigmoidoscopy (FS) were invited between November 2006 and November 2007, for once-only FS. Those allocated to colonoscopy were invited between June 2009 and August 2010 for once-only colonoscopy.

Symptomatic persons with rectal blood loss or a recent change in bowel habits were asked not to participate but to consult their general practitioner. Persons with a history of CRC or inflammatory bowel disease were also asked not to participate. Participants reporting a colonoscopy or CT-colonography in the past 3 years, those with an estimated life expectancy of less than 5 years, and those who were not able to give informed consent were excluded. In the FIT-screening cohort,

persons were not re-invited for following rounds in case of a positive FIT in a previous round, passing the upper age limit or, or moving out of the selected postal area.

Intervention

All invitees received an advance notification letter, followed by a kit two weeks later by postal mail. This kit contained an invitation letter, information brochure, and an informed consent form. For FIT-invitees the kit further contained a single FIT-test and testing instructions. For FS-invitees and colonoscopy-invitees the invitation letter contained a telephone number of the screening unit to schedule an appointment for an intake. A reminder was sent six weeks after the initial invitation to invitees who had not yet responded.

Fecal immunochemical test

Persons allocated to FIT received one FIT every two years. They were instructed to collect a single sample of one bowel movement. In the first 3 rounds invitees received the OC-Sensor (Eiken, Japan), and in the fourth round invitees were randomized to receive either the OC-Sensor or the FOB-Gold (Sentinel, Italy). Invitees were asked to sample feces according to instructions and post the feces sample within 24 hours after collection, keeping the sample in the refrigerator until mailing. Participants signed an informed consent form with the date of sample collection. The test result was considered positive when the hemoglobin concentration in the FIT sample was ≥10 µg Hb/g faeces. Previous results have shown that the OC-Sensor and FOB-Gold perform equally well over the relevant concentration range [23].

Flexible sigmoidoscopy

Invitees allocated to an intake appointment who had scheduled for sigmoidoscopy received a phosphate enema by mail with instructions for self-administration. Administration of the enema by a nurse in the screening unit was offered as an alternative. All sigmoidoscopies were performed by experienced endoscopists in a dedicated centre. The FS was defined as complete when reaching the splenic flexure. Participants did not receive sedatives. Participants were referred for colonoscopy when one of the following criteria was met: presence of a polyp with a diameter ≥10 mm; an adenoma with serrated, villous histology (≥25% villous) or high-grade dysplasia; ≥3 adenomas; ≥20 hyperplastic polyps; or invasive CRC.

Colonoscopy

All FIT-positive, FS-positive or once-only colonoscopy participants were scheduled for colonoscopy,

unless contraindicated. Experienced endoscopists (>1000 colonoscopies) performed all colonoscopies for this study. Participants received standard bowel preparation: low-fibre diet and oral intake of 2 litres of transparent fluid and a laxative solution (Moviprep; Norgine, Amsterdam, The Netherlands or Picoprep; Ferring, Hoofddorp, The Netherlands) at home. Colonoscopy was performed under conscious sedation with carbondioxide insufflation. All quality parameters were recorded on a standardized clinical record form (CRF). Polyps were removed during the same procedure if possible. If immediate endoscopic treatment was impossible, the screenee was either planned for a second look colonoscopy to remove the lesion or, biopsies were obtained and pathological assessment of these tissue samples provided a definitive diagnosis and participant was referred to our colorectal multidisciplinary team for further treatment. In case of incomplete colonoscopy, a computed tomographic colonography (CTC) was performed. All complications were registered. Surveillance after removal of adenomatous polyps, large (≥10 mm) serrated lesions or cancer was recommended according to the Dutch surveillance guideline[24]. Screenees with a negative colonoscopy were discharged from screening for 10 years.

Colorectal Lesions

The following data on colorectal lesions were collected with the case report form: location, size, macroscopic aspect, and morphology. Additionally, details on polypectomy technique and endoscopic assessment of radicality were recorded for all lesions detected during colonoscopy. All removed lesions were collected and evaluated by an experienced gastrointestinal pathologist according to the Vienna criteria[25]. The lesions were classified as adenoma (tubular, tubulovillous, villous), serrated polyp (hyperplastic, sessile serrated adenoma, traditional serrated adenoma), adenocarcinoma or miscellaneous. Dysplasia was defined as either low-grade or high-grade. Advanced adenomas were defined as adenomas ≥ 10 mm, adenomas with high-grade dysplasia or adenomas with a villous component of at least 25%. Cancers were staged according to the American Joint Committee on Cancer classification[26]. AN was defined as advanced adenoma and/or colorectal cancer.

Screen detected and Interval colorectal cancers

Screen detected CRCs were defined as CRCs detected by screening. An interval colorectal cancer was defined as a CRC diagnosed in the interval between a negative screen and the next recommended exam[27]. Data from all invitees were linked to our Dutch Comprehensive Cancer Centre to identify interval cancers, which was up to date until March 2015.

Statistical analysis

The main outcome of this study was the diagnostic yield of AN. This was analysed both in an intention-to-screen analysis, where it was defined as the number of screenees with AN relative to all invitees, and as an as-screened analysis, in which it was defined as the number of screenees with AN relative to all participants. Additionally, we evaluated participation rate, positivity rate and colonoscopy rate for each screening strategy. For FIT-screening cumulative rates over four rounds were used in these analyses. Moreover, results per FIT-screening round were separately calculated and presented. Participation rate was calculated as the number of invitees returning a FIT relative to the number of all eligible invitees. Positivity rate was defined as the proportion of participants having a positive test result relative to the number of tests returned. Differences in means were analysed using Student's t-test. Differences in proportions were analysed using Chi-square testing. Participation rate, positivity rate and diagnostic yield are reported as proportions with 95% confidence intervals. All tests were conducted using SPSS version 21.0 (SPSS Inc, Chicago, Ill).

Ethical approval

The Dutch National Health Council approved the study (decisions 2006/02 WBO, 2009/03WBO, 2013/20 WBO, The Hague, Netherlands). All screenees gave written informed consent.

Results

Screening population

A total of 30,052 average-risk persons were randomly selected to be invited for colorectal cancer screening. Five percent of these (n=1,485) did not meet the inclusion criteria, resulting in 14,651 invitees for FIT-screening, 7,882 invitees for FS-screening, and 5,982 invitees for colonoscopy screening. The median age was similar for all three screening groups: 59 years for FIT (IQR 55-65 years), 59 years for FS (IQR 54-65) and 60 years for colonoscopy (IQR 54-65). There were no gender differences, with 50% (n=14,328) of the invitees being male (FIT n=7,264 (50%); FS n=3,941 (50%); colonoscopy n= 2,982 (50%)).

Participation and colonoscopy rates

An overview of the study design, participation rates and adherence to diagnostic follow-up is provided in Figure 1. In the FIT-group, participation rates were significantly higher with 73% of invitees participating at least once and 60% participating to the first invitation, compared to 30.9% of invitees in the FS group and 23.8% of invitees in the colonoscopy group ($p<0.001$). Among FIT-

screenees, 18.9% had a positive FIT in one of the four rounds, whereas 9% of FS-screenees had a positive test. As a result of these participation and positivity rates, a total of 12.8% of those invited for FIT-screening underwent a colonoscopy, for the FS group this figure was 2.7% and for the colonoscopy group this was the highest with 23.8% (p<0.001).

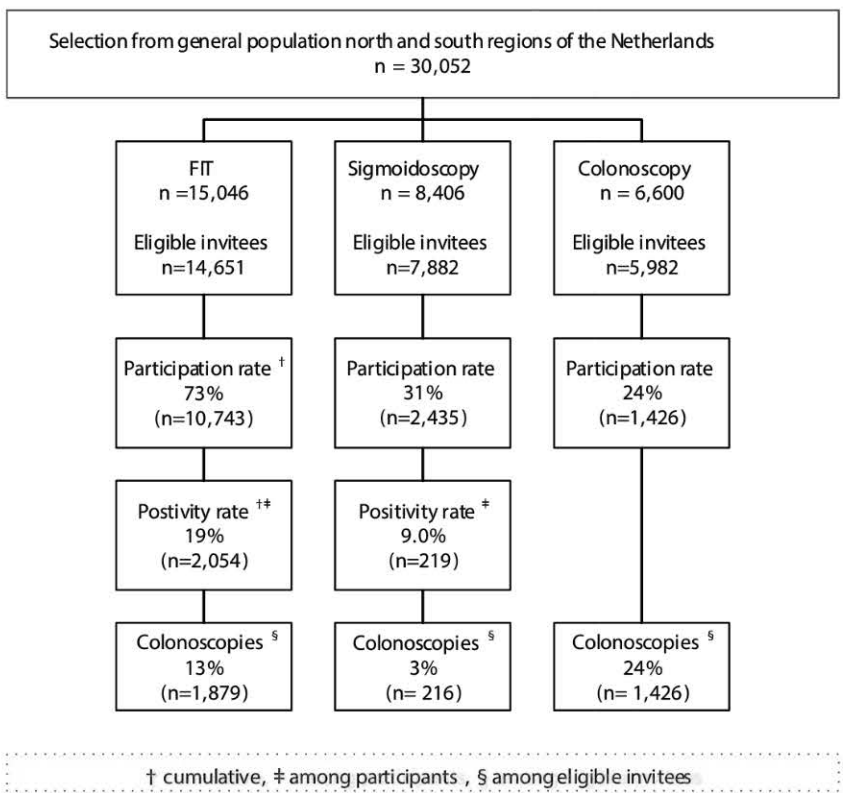


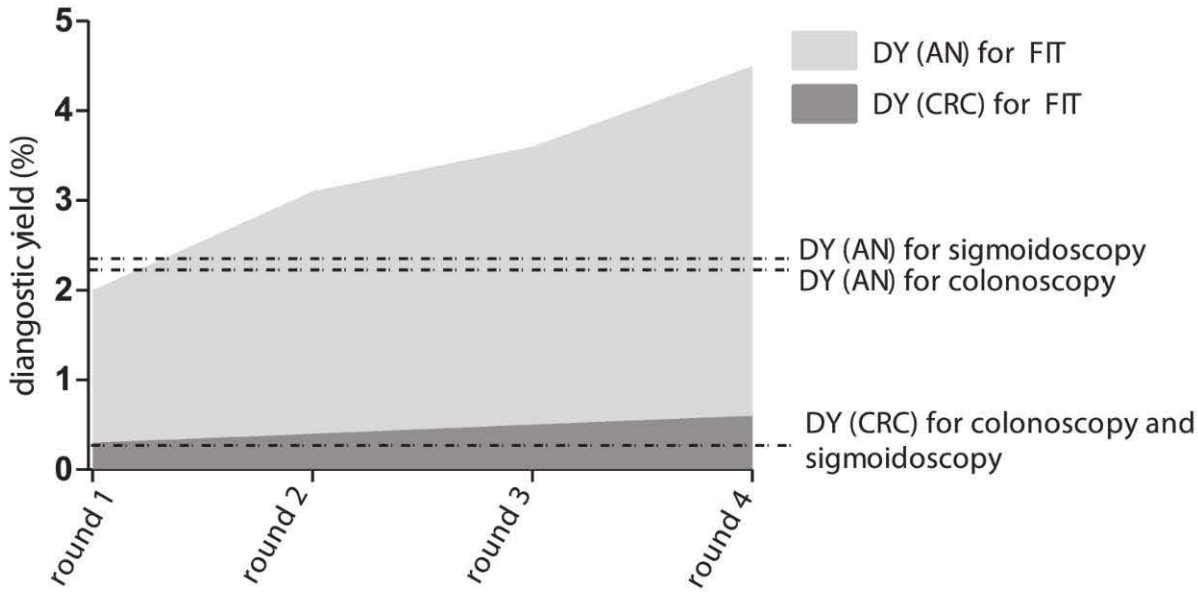
Figure 1. *Study flow: three screening trials.*

Diagnostic yield

In the intention-to-screen analysis including all invitees, FIT detected significantly more AN compared to FS and colonoscopy; 4.5% (95% CI 4.2-4.9) versus 2.3% (95% CI 2.0-2.7) and 2.2% (95% CI 1.8-2.6; Table 1), respectively. Moreover, FIT detected three times more CRC per invitee than endoscopic screening: 0.6% (95% CI 0.5-0.7) versus 0.2% (95% CI 0.1-0.3; Table 1) for both colonoscopy and FS. Higher rates of non-advanced adenomas and small serrated polyps were found for colonoscopy and FS compared to FIT. Among screenees with serrated polyps, similar rates of sessile serrated adenomas or large hyperplastic polyps (i.e. >10mm) were found for all strategies, with 0.2% (95% CI 0.1-0.3) for FIT, 0.1% (95% CI 0.1-0.2) for FS and 0.3% (95% CI 0.3-0.5) for colonoscopy. Figure 2 shows the cumulative increase in diagnostic yield for AN and CRC (supplementary Table 1 provides the cumulative participation rate, positivity rate, colonoscopy rate and diagnostic yield among invitees

of FIT-screening over four rounds).

A.



B.

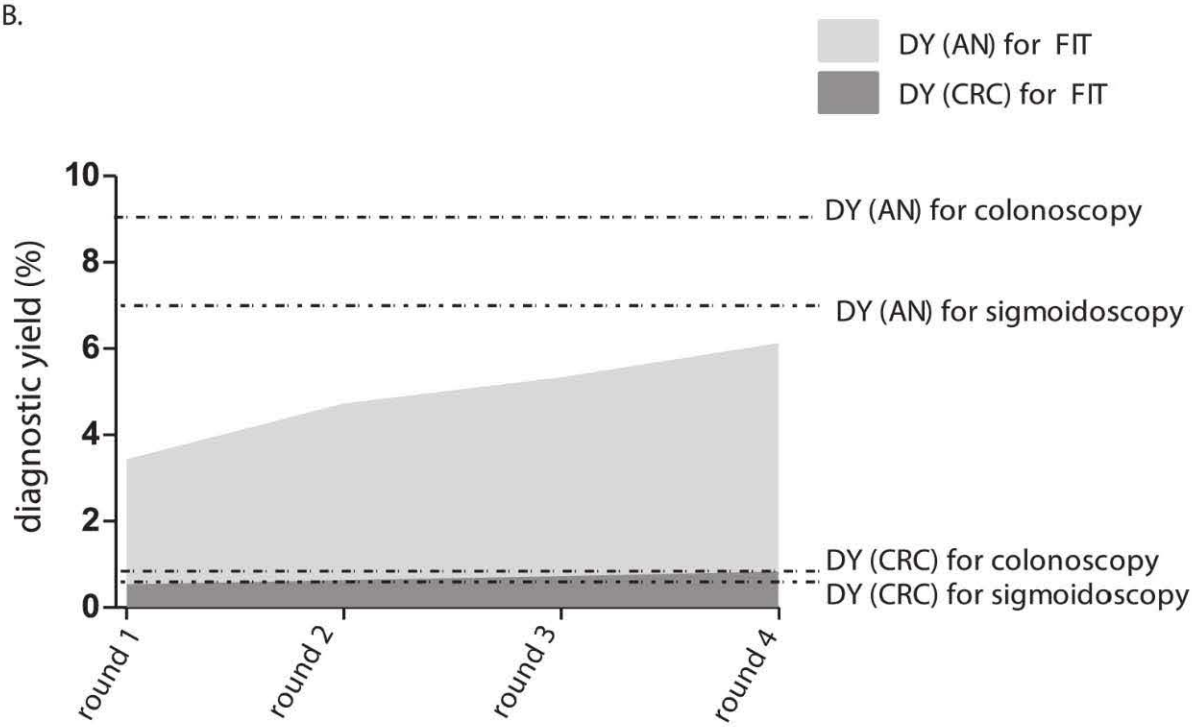


Figure 2. *Cumulative diagnostic yield (DY) over four rounds of FIT-screening for advanced neoplasia (AN) and colorectal cancer (CRC) in the intention-to-screen analysis* **(A)** *and as-screened analysis* **(B)**.

Table 1. *Diagnostic yield of three screening strategies.*

	FIT	Sigmoidoscopy	Colonoscopy
	% (95% CI)	% (95% CI)	% (95% CI)
Intention-to-screen	(n=14,651)	(n=7,882)	(n=5,982)
Advanced neoplasia	4.5 (4.2-4.9)	2.3 (2.0-2.7)	2.2 (1.8-2.6)
Colorectal cancer	0.6 (0.5-0.7)	0.2 (0.1-0.3)	0.2 (0.1-0.3)
Advanced adenomas	3.9 (3.6-4.3)	2.1 (1.8-2.4)	2.0 (1.7-2.4)
Non-advanced adenomas	3.2 (2.9-3.5)	3.7 (3.3-4.1)	5.6 (5.0-6.2)
Serrated polyps	1.1 (0.9-1.3)	5.9 (5.4-6.4)	3.9 (3.5-4.5)
As-screened	(n=10,743)	(n=2,435)	(n=1,426)
Advanced neoplasia	6.1 (5.7-6.6)	7.4 (6.5-8.5)	9.1 (7.7-10.7)
Colorectal cancer	0.8 (0.6-0.9)	0.5 (0.3-0.9)	0.6 (0.3-1.2)
Advanced adenomas	5.4 (5.0-5.8)	6.7 (5.8-7.8)	8.5 (7.1-10.0)
Non-advanced adenomas	4.3 (3.9-4.7)	12.0 (10.7-13.3)	23.4 (21.3-25.7)
Serrated polyps	1.5 (1.3-1.7)	19.1 (17.5-20.7)	16.5 (14.6-18.5)
Non-screen detected *	(n=14,651)	(n=7,882)	(n=5,982)
Interval cancer	0.2 (0.1-0.3)**	0.1 (0-0.2)***	0.01 (0-0.1)

* maximum follow-up: colonoscopy= 5.8 years , FS= 5 years, FIT = 8.3 years
** 0.13 % FIT-interval cancer 0.05% post-colonoscopy interval cancer
** 0.09 % FS-interval cancer 0.03% post-colonoscopy interval cancer

Interval cancers

Linkage to the Dutch Comprehensive Cancer Centre provided data on interval cancers (Table 1). Among invitees, 19 (0.13%) FIT-negative screenees developed a cancer within the screening interval, 6 (0.09%) FS-negative screenees, and 1 (0.01%) colonoscopy screenee developed CRC, despite the absence of advanced neoplasia at colonoscopy. Furthermore, 7 post-colonoscopy CRC (0.05%) were diagnosed in those with a positive FIT, and 1 (0.03%) in a participant with a positive-FS. Taking into account these program-related interval cancers, endoscopic screening had a significantly lower interval cancer rate compared to FIT-screening (Table 1). However, it should be noted that colonoscopy had a much shorter follow-up time than FIT-screening, 5.8 years compared to 8.3 years respectively.

In the as-screened analysis, including only those who participated in screening, colonoscopy detected more AN (9.1%; 95% CI 7.7-10.7) than FS (7.4%; 95% CI 6.5-8.5) and FIT (6.1%; 95% CI 5.7-6.6). However, CRC detection rates were similar for FS (0.5%; 95% CI 0.3-0.9), colonoscopy (0.6%; 95% CI 0.3-1.2) and FIT (0.8; 95% CI 0.6-0.9). Colonoscopy detected more non-AN, with almost a quarter of the participants having non-AN as most advanced finding (23.4; 95% CI 21.3-25.7).

Regarding serrated polyps, FIT-screening yielded the lowest rate of 1.5% (95% CI 1.3-1.7) compared to colonoscopy (16.5%; 95% CI 14.6-18.5) and FS (19.1%; 95% CI 17.5-20.7). No differences were

found regarding stage and location of CRCs among the three screening strategies (Table 2).
The number needed to invite (NNI) to detect one participant with AN was 22 for FIT, 43 for FS and 46 for colonoscopy. The NNI to detect one CRC was 178, 606 and 664, respectively. [NNS toevoegen] Next, the number needed to scope by colonoscopy (NNSc) to detect one AN was 2 for FIT, 3 for FS and 11 for colonoscopy. The NNSc to detect one CRC was 23, 17 and 159, respectively.

Table 2. *Location and stages of screen-detected cancer and interval cancers.*

	FIT		FS		Colonoscopy	
	Screen-detected cancer	Interval cancer	Screen-detected cancer	Interval cancer	Screen-detected cancer	Interval cancer
	% (n=83*)	% (n=19*)	% (n=13)	% (n=10*)	% (n=9)	% (n=1)
Stage						
I	54 (45)	26 (5)	77 (10)	20 (2)	78 (7)	-
II	13 (11)	16 (3)	0 (0)	20 (2)	11 (1)	100 (1)
III	32 (26)	37 (7)	23 (3)	40 (4)	11 (1)	-
IV	1 (1)	21 (4)	0 (0)	20 (2)	0 (0)	-
Location						
Distal	68 (56)	58 (11)	85 (11)	3 (30)	44 (4)	-
Proximal	32 (27)	42 (8)	15 (2)	7 (70)	56 (5)	100 (1)

*post-colonoscopy CRC not included

Stages and location of screen-detected and interval cancers

Location and stages of screen-detected cancers and interval cancers are described in Table 2. No significant differences were seen between the three screening strategies when comparing stages of screen-detected cancers (p=0.54), nor between location of lesions (i.e. distal vs. proximal; p=0.19). However, it should be noted that few proximal cancers were detected in the FS-screening group. Regarding the interval cancers after a negative FIT or FS, participants had a comparable distribution of cancer stage, both with around 20% stage IV cancers. Notably, most interval cancers in the FS group were located in the proximal colon. Only one interval cancer was detected in the colonoscopy-arm, which was a stage II tumor and located in the proximal colon.

Discussion

This analysis of three large CRC screening trials is the first to report outcomes of endoscopic-screening versus multiple rounds of FIT-screening in an average-risk population. Our results indicate a higher diagnostic yield of AN among invitees in FIT-screening, compared to FS-screening and colonoscopy-screening. No differences were found regarding the detection of CRC among all three

strategies. This yield was reached with significantly fewer colonoscopies for FIT- and FS-screening compared to colonoscopy-screening. Among participants, colonoscopy had the highest diagnostic yield for AN, while also detecting the largest number of clinically non-relevant neoplasia.

Our study has several strengths; firstly, invitees were all randomly selected from the same population, using equal criteria for selection. All data were prospectively collected and screenees were linked to the Netherlands Cancer Registry by the Dutch Comprehensive Cancer Centre to account for interval cancers. Since 1989, the Netherlands Cancer Registry registers all patients diagnosed with cancer in the Netherlands and provides a unique and fully covered database. Furthermore, for all three screenings strategies, all endoscopies were performed in the same endoscopy centres with similar quality indicators and performance. As there was no national screening program at the start of the screening studies, this could not have influenced awareness and subsequently participation rates.
To fully appreciate our results, some limitations also need to be addressed. The colonoscopy-screening trial started three years later than the FIT-screening and FS-screening trials, leading to different follow-up times. This makes comparing interval cancer rate between the three modalities arduous. Additionally, ideally one more round of FIT and subsequent two-year follow-up should be completed to encompass the same time frame (i.e. 10 years) for each screening modality and thus to be able to fully compare the results to colonoscopy screening.

Participation rates are of crucial importance in a screening program, as they directly affect diagnostic yield. While a small difference in uptake might seem irrelevant, in population-based screening such small percentages can lead to large numbers of additional screenees[28]. In our study, participation rates were highest for FIT and lowest for colonoscopy screening. Though participation rates vary widely geographically, endoscopic screening consistently shows lower participation rates compared to FOBT-based screening[3, 4, 10, 29-31]. A previously published randomized controlled study in Italy showed participation rates after one round of screening of 14.8% for colonoscopy and 50.4% for FIT-screening[29]. A Spanish randomized controlled study comparing colonoscopy and FIT-screening, showed participation rates after one round of screening of 24.6% and 34.2%, respectively. Notably, in both studies FIT participation rates were substantially lower than in our study and study recruitment strategies differed from our study. Other studies have reported participation rates varying from 46% to 63% over multiple rounds of FIT-screening, with up to 78% of FIT screenees attending screening at least once over four rounds of screening[32, 33]. These latter findings are in line with our results, with around 73% of FIT-screenees participating at least once. Our findings also confirm the stable attendance rates over multiple rounds of FIT, as described in other large screening cohorts[32, 33].

In the intention-to-screen analysis, which comprises the total eligible study group, diagnostic yield was highest for FIT-screening. FIT-screening in particular identified three-times more participants with CRC than endoscopic screening. Though colonoscopy yielded significantly more screenees with AN in the as-screened analysis, CRC detection rates for participants were similar for all three strategies. After four rounds of FIT-screening, the diagnostic yield of CRC was even slightly higher with FIT than after once-only colonoscopy. A fifth FIT-screening round would likely increase the difference and thus potentially reach the level of statistical significance. However, comparing CRC detection rates of FIT and endoscopic screening is complex and results should be interpreted with caution, as CRCs detected in FIT-screening could in theory have been prevented in a once-only colonoscopy by the removal of advanced adenomas.

Comparing results between screening programs in different countries is also challenging and should be done cautiously, because of differences in organizational structures, populations, and quality parameters.(34) The previously mentioned Spanish study found that FIT and colonoscopy yielded similar CRC detection rates even after only one round of screening.(4) Future screening rounds of that study will determine whether this effect will increase by time and what the influence is on the interval cancer rate. In some settings FS is repeated every 5 years, for these settings diagnostic yield of once-only FS over 5 years might best be compared to three rounds of FIT-screening. Our results showed a higher diagnostic yield for FIT both for AN and CRC over three rounds, with similar interval cancer rates.

CRC screening puts a large demand on colonoscopy capacity, and consequently many countries struggle with a gap between available and required resources. In some other countries, like the United States, capacity is less of a problem, but high use of an expensive modality strains budgets. Moreover, as it is an invasive procedure it comes with considerable risks and burden for patients. FIT and FS identify participants at higher-than-average risk for AN to selectively refer those for colonoscopy. As a result, the number needed to scope to detect AN in one screenee was 2 for FIT-screening, 3 for FS, compared to 11 for once-only colonoscopy. FIT and FS screening require substantially fewer colonoscopies to yield a similar diagnostic yield.

The low detection rate of non-AN and serrated polyps with FIT could be seen as an advantage, as it would reduce the number of screenees that would have to undergo an unnecessary colonoscopy, (i.e., a negative colonoscopy) which in turn would have a beneficial influence on colonoscopy-related resources and costs. On the other hand, the lower use of colonoscopy also implies that

lesions remain undetected and may progress to cancer. This in particular pertains to sessile serrated adenomas (SSA). Our data show that for all three screening strategies similar rates of SSA as a most advanced finding are detected and that these were found in only a small number of participants. A recent study has shown that sessile serrated adenomas are associated with the existence of synchronous AN. This would suggest that FIT would possibly also detect these SSA's through the detection of occult blood from the also present AN [35].

In conclusion, different screening strategies are associated with marked differences in uptake and diagnostic yield. Over four rounds, FIT was the most effective strategy in population-based CRC screening, leading to detection of high rates of CRC and AN, while requiring the lowest colonoscopy demand. Since many countries are considering to implement screening programs, the findings of this study aid in deciding on choice of screening strategy worldwide, based on expected participation rates and available colonoscopy resources.

References

1. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. Gut. 2016:[E-pub ahead of print] doi:10.1136/gutjnl-2015-310912.
2. Kuipers EJ, Rosch T, Bretthauer M. Colorectal cancer screening--optimizing current strategies and new directions. Nat Rev Clin Oncol. 2013;10(3):130-42.
3. 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.
4. Quintero E, Castells A, Bujanda L, Cubiella J, Salas D, Lanas A, et al. Colonoscopy versus fecal immunochemical testing in colorectal-cancer screening. N Engl J Med. 2012;366(8):697-706.
5. Kaminski MF, Bretthauer M, Zauber AG, Kuipers EJ, Adami HO, van Ballegooijen M, et al. The NordICC Study: rationale and design of a randomized trial on colonoscopy screening for colorectal cancer. Endoscopy. 2012;44(7):695-702.
6. Colonoscopy Versus Fecal Immunochemical Test in Reducing Mortality From Colorectal Cancer (CONFIRM) (CONFIRM): ClinicalTrials.gov; 2010 [updated 2016; cited 2016]. Available from: https://clinicaltrials.gov/ct2/show/NCT01239082.
7. Brenner H, Stock C, Hoffmeister M. Effect of screening sigmoidoscopy and screening colonoscopy on colorectal cancer incidence and mortality: systematic review and meta-analysis of randomised controlled trials and observational studies. BMJ. 2014;348:g2467.
8. 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.
9. 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.
10. Segnan N, Senore C, Andreoni B, Azzoni A, Bisanti L, Cardelli A, et al. Comparing attendance and detection rate of colonoscopy with sigmoidoscopy and FIT for colorectal cancer screening. Gastroenterology. 2007;132(7):2304-12.
11. Shaukat A, Mongin SJ, Geisser MS, Lederle FA, Bond JH, Mandel JS, et al. Long-term mortality after screening for colorectal cancer. N Engl J Med. 2013;369(12):1106-14.
12. 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.
13. Brenner H, Tao S. Superior diagnostic performance of faecal immunochemical tests for haemoglobin in a head-to-head comparison with guaiac based faecal occult blood test among 2235 participants of screening colonoscopy. Eur J Cancer. 2013.
14. Faivre J, Dancourt V, Denis B, Dorval E, Piette C, Perrin P, et al. Comparison between a guaiac and three immunochemical faecal occult blood tests in screening for colorectal cancer. European Journal of Cancer. 2012;48(16):2969-76.
15. 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.
16. Sung JJ, Ng SC, Chan FK, Chiu HM, Kim HS, Matsuda T, et al. An updated Asia Pacific Consensus Recommendations on colorectal cancer screening. Gut. 2015;64(1):121-32.
17. 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. Am J Gastroenterol. 2012;107(10):1570-8.
18. Gellad ZF, Stechuchak KM, Fisher DA, Olsen MK, McDuffie JR, Ostbye T, et al. Longitudinal adherence to fecal occult blood

testing impacts colorectal cancer screening quality. Am J Gastroenterol. 2011;106(6):1125-34.

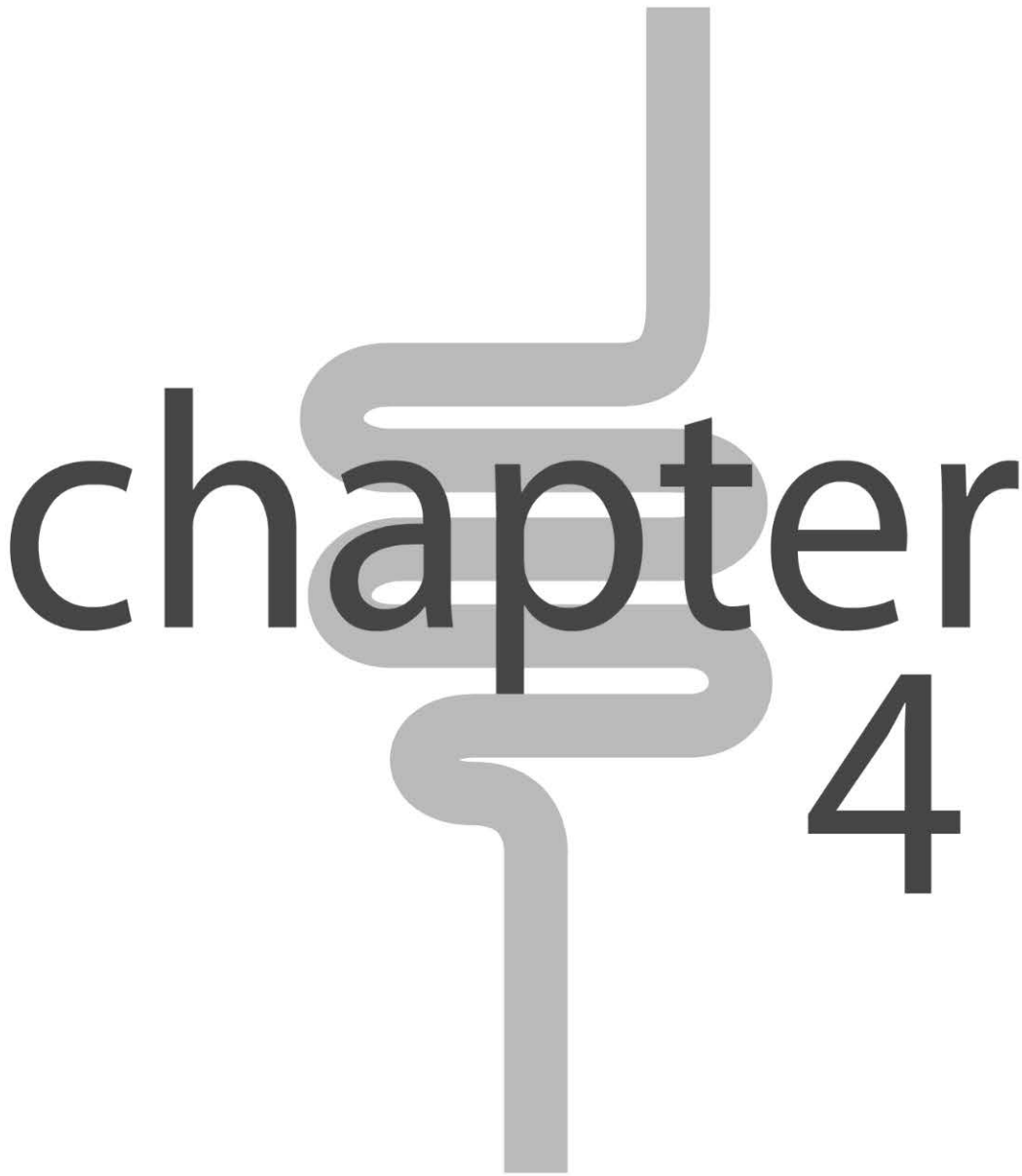
19. Kapidzic A, Grobbee EJ, Hol L, van Roon AH, van Vuuren AJ, Spijker W, et al. Attendance and yield over three rounds of population-based fecal immunochemical test screening. Am J Gastroenterol. 2014;109(8):1257-64.
20. de Wijkerslooth TR, de Haan MC, Stoop EM, Deutekom M, Fockens P, Bossuyt PM, et al. Study protocol: population screening for colorectal cancer by colonoscopy or CT colonography: a randomized controlled trial. BMC Gastroenterol. 2010;10:47.
21. Stegeman I, van Doorn SC, Mundt MW, Mallant-Hent RC, Bongers E, Elferink MA, et al. Participation, yield, and interval carcinomas in three rounds of biennial FIT-based colorectal cancer screening. Cancer Epidemiol. 2015;39(3):388-93.
22. Hol L, van Leerdam ME, van Ballegooijen M, van Vuuren AJ, van Dekken H, Reijerink JC, et al. Screening for colorectal cancer: randomised trial comparing guaiac-based and immunochemical faecal occult blood testing and flexible sigmoidoscopy. Gut. 2010;59(1):62-8.
23. Grobbee EJ, van der Vlugt M, van Vuuren A, Stroobants AK, Didden P, Mundt MW, et al. Comparison of OC-Sensor and FOB-Gold in Population-Based Colorectal Cancer Screening Based on FIT. Gastroenterology. 2016;148(4):S-160.
24. Dekker E, van Leerdam ME, Hazewinkel Y, Sanduleanu S, Vasen HF, Lansdorp-Vogelaar I, et al. Nederlandse Richtijn Coloscopie Surveillance: Nederlandse Vereniging voor Maag-, Darm- en Leverziekten; 2013. Available from: http://www.mdl.nl/uploads/240/1308/Richtlijn_Coloscopie_Surveillance_definitief_2013.pdf.
25. Schlemper RJ, Riddell RH, Kato Y, Borchard F, Cooper HS, Dawsey SM, et al. The Vienna classification of gastrointestinal epithelial neoplasia. Gut. 2000;47(2):251-5.
26. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. Ann Surg Oncol. 2010;17(6):1471-4.
27. 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.
28. Kuipers EJ, Spaander MC. Colorectal Cancer Screening by Colonoscopy, CT-Colonography, or Fecal Immunochemical Test. J Natl Cancer Inst. 2016;108(2).
29. Sali L, Mascalchi M, Falchini M, Ventura L, Carozzi F, Castiglione G, et al. Reduced and Full-Preparation CT Colonography, Fecal Immunochemical Test, and Colonoscopy for Population Screening of Colorectal Cancer: A Randomized Trial. J Natl Cancer Inst. 2016;108(2).
30. Gupta S, Halm EA, Rockey DC, Hammons M, Koch M, Carter E, et al. Comparative effectiveness of fecal immunochemical test outreach, colonoscopy outreach, and usual care for boosting colorectal cancer screening among the underserved: a randomized clinical trial. JAMA Intern Med. 2013;173(18):1725-32.
31. Inadomi JM, Vijan S, Janz NK, Fagerlin A, Thomas JP, Lin YV, et al. Adherence to colorectal cancer screening: a randomized clinical trial of competing strategies. Arch Intern Med. 2012;172(7):575-82.
32. Crotta S, Segnan N, Paganin S, Dagnes B, Rosset R, Senore C. High rate of advanced adenoma detection in 4 rounds of colorectal cancer screening with the fecal immunochemical test. Clin Gastroenterol Hepatol. 2012;10(6):633-8.
33. Jensen CD, Corley DA, Quinn VP, Doubeni CA, Zauber AG, Lee JK, et al. Fecal Immunochemical Test Program Performance Over 4 Rounds of Annual Screening: A Retrospective Cohort Study. Ann Intern Med. 2016;164(7):456-63.
34. Klabunde C, Blom J, Bulliard JL, Garcia M, Hagoel L, Mai V, et al. Participation rates for organized colorectal cancer screening programmes: an international comparison. J Med Screen. 2015;22(3):119-26.
35. IJspeert JE, de Wit K, van der Vlugt M, Bastiaansen BA, Fockens P, Dekker E. Prevalence, distribution and risk of sessile serrated adenomas/polyps at a center with a high adenoma detection rate and experienced pathologists. Endoscopy. 2016.

Supplementary file

Supplemantary table 1. *Cumulative participation rate, positivity rate, colonoscopy rate and diagnostic yield among invitees of FIT-screening over four rounds (n=14,651)*

	2 rounds	3 rounds	4 rounds
	% (95% CI)	% (95% CI)	% (95% CI)
Participation rate	67 (66-68)	71 (70-72)	73 (73-74)
Positivity rate*	12 (12-13)	16 (15-16)	19 (18-20)
Colonoscopy rate	8 (7-8)	10 (10-11)	13 12-13)
Advanced neoplasia	3.1 (2.8-3.4)	3.7 (3.4-4.0)	4.5 (4.2-4.9)
Colorectal cancer	0.4 (0.3-0.5)	0.5 (0.4-0.6)	0.6 (0.5-0.7)
Non-advanced adenomas	1.7 (1.5-1.9)	2.4 (2.2-2.7)	3.2 (2.9-3.5)
Serrated polyps	0.6 (0.5-0.7)	0.8 (0.7-1.0)	1.1 (0.9-1.3)

*among participants


chapter
4

A randomized comparison of two fecal immunochemical tests in population-based colorectal cancer screening

Esmée J. Grobbee, Manon van der Vlugt, Anneke J. van Vuuren, An K. Stroobants, Marco W. Mundt, Wolfert J. Spijker, Evelien .J.C. Bongers, Ernst J. Kuipers, Iris Lansdorp-Vogelaar, Patrick M. Bossuyt, Evelien Dekker, Manon C.W. Spaander

Gut, 2016

Abstract

Objective Colorectal cancer (CRC) screening programs are implemented worldwide; many are based on fecal immunochemical testing (FIT). The aim of this study was to evaluate two frequently used FITs on participation, usability, positivity rate and diagnostic yield in population-based FIT-screening.

Design Comparison of two FITs was performed in a fourth round population-based FIT-screening cohort. Randomly selected individuals aged 50-74 were invited for FIT-screening and were randomly allocated to receive an OC-sensor (Eiken, Japan) or FOB-Gold (Sentinel, Italy) test (March-December 2014). A cut-off of 10 µg Hb/g feces (i.e. 50 ng Hb/ml buffer for OC-sensor and 59 ng Hb for FOB-Gold) was used for both FITs.

Results In total 19,291 eligible invitees were included (median age 61, IQR 57-67; 48% males): 9,669 invitees received OC-Sensor and 9,622 FOB-Gold; both tests were returned by 63% of invitees ($p=0.96$). Tests were non-analysable in 0.7% of participants using OC-Sensor vs. 2.0% using FOB-Gold ($p<0.001$). Positivity rate was 7.9% for OC-Sensor, and 6.5% for FOB-Gold ($p=0.002$). There was no significant difference in diagnostic yield of advanced neoplasia (1.4% for OC-Sensor vs. 1.2% for FOB-Gold; $p=0.15$) or positive predictive value (PPV; 31% versus 32%; $p=0.80$). When comparing both tests at the same positivity rate instead of cut-off, they yielded similar PPV and detection rates.

Conclusion The OC-Sensor and FOB-Gold were equally acceptable to a screening population. However, FOB-Gold was prone to more non-analysable tests. Comparison between FIT brands is usually done at the same Hb stool concentration. Our findings imply that for a fair comparison on diagnostic yield between FITs positivity rate rather than Hb-concentration should be used.

Introduction

Colorectal cancer (CRC) is one of the major causes of death in the Western World[1,2,3]. Population based colorectal cancer screening aims to detect CRC and its precursors in an earlier phase, thereby reducing CRC morbidity and mortality[4]. Of the currently available screening tests, fecal occult blood tests (FOBT) and sigmoidoscopy are the only strategies that have been proven in prospective randomized controlled trials to reduce CRC-related mortality[5]. The evidence favouring fecal immunochemical testing (FIT) over guaiac fecal occult blood testing (gFOBT) is substantial[6, 7, 8, 9, 10, 11]. FIT is more likely to detect hemoglobin from the lower GI-tract than most gFOBTs. FIT is also easier to use resulting in higher participation rates. In addition, FIT enables quantitative measuring fecal hemoglobin (Hb), allowing the use of different cut-off concentrations, and can be analyzed by automation.

Many countries have implemented FIT-based CRC screening programs or are about to do so[12]. When implementing a national screening program, choosing the appropriate test is of key importance. FIT tests can be based on dry sampling, using cards, or wet sampling, using tubes with buffers[8]. The most frequently used FITs involve wet sampling (i.e. storage and transport of feces in a wet preservative), and different FIT-brands in this category are available. These various FIT-brands often have different sampling tubes and buffer volumes, resulting in different expressions of Hb-concentration that are not interchangeable[13]. Such differences complicate direct comparison of FITs. It has been proposed to standardize quantitative FIT results in µg Hb per gram feces, allowing diagnostic test accuracy to be compared more easily between FITs using the same standardized cut-off[14]. At present, there is no evidence favouring a specific FIT brand[15].

As small differences in test characteristics can have major effects on a population-level, it is important to further assess brand-related differences. In the Netherlands, several pilot-studies have been performed using the OC-Sensor (Eiken, Japan)[16, 17, 18]. However, the recently started nationwide program has selected the FOB-Gold test (Sentinel, Italy) through a European bid. Few comparative data are available for these two tests. To the best of our knowledge, no study has investigated the OC-Sensor and FOB-Gold head-to-head in a screening setting. Such a comparison is relevant since both tests are among the most widely used FIT tests. We therefore aimed to compare the OC-Sensor and FOB-Gold with regard to participation rate, usability, positivity rate and diagnostic yield, using a standardized cut-off of fecal Hb concentration.

Methods

Study population/study design

Details about the design of the on-going population-based CRC screening pilot-program have been described previously[7, 16, 17, 19, 20]. This trial was registered at www.trialregister.nl (identifier NTR5385). In short, from June 2006 the demographic data of all individuals between 50 and 74 years living in a selected region in the southwest and northwest of the Netherlands were obtained from municipal population registers. As there was no CRC screening program at the time of the trial the target population was screening-naive when first approached.

This study was part of a dynamic (i.e., including all subjects included in any of the first three rounds as well as individuals who moved into the target areas and those who had reached the target age) cohort study,. Individuals who had moved out of the selected area or passed the upper-age limit were not re-invited. Invitations for this 4th screening round were done similar to previous rounds. This implied that for the Southern region, random samples were taken from the target population by a computer-generated algorithm (Tenalea, Amsterdam, The Netherlands). In the Northern region random samples of selected postal code areas were taken. For this fourth round of screening both cohorts were combined. Invitations were sent out between March 2014 and December 2014. Individuals with a history of inflammatory bowel disease or CRC, as well as those who had undergone a colonoscopy in the past 2 years, had an estimated life expectancy of <5 years, or were unable to give informed consent were excluded from the study.

Intervention and randomization; two types of FIT-screening

All invitees were randomly allocated to receive either the OC-sensor (Eiken, Japan) or the FOB-Gold (Sentinel, Italy) sorted according to household without stratification. Randomization was performed before invitation in a 1:1 ratio. People in the same household were allocated to the same test, to avoid confusion in handling the FIT. The FIT was sent by mail with the instructions to collect a single sample of one bowel movement using a collection device probe. The test result was considered positive when the hemoglobin concentration in the FIT sample was ≥10 µg Hb/g feces. This cut-off corresponds to 50 ng Hb/ml for the OC-sensor and 59 ng Hb/ml for the FOB-Gold. In case of an non-analysable test a new FIT was sent to the participant.

FIT analysis

The OC-sensor device collected 10 mg feces with a serrated shaped probe attached to the cap in

2.0 ml preservative buffer. The FOB-gold pierceable tube collected 10 mg feces with a serrated shaped probe attached to the cap in 1.7 ml preservative buffer. Participants were asked to dip the probe four times in the feces and to reinsert the probe into the respective device. Participants were asked to post the feces samples within 24 hours after collection and keep the sample in the refrigerator. Together with the FIT test participants signed an informed consent form with the date of sample collection. Participants returned FIT test and permission form at ambient temperature by freepost to two laboratories (Laboratory Clinic Chemistry, Academic Medical Centre, Amsterdam or Gastroenterology & Hepatology laboratory, Erasmus Medical Centre, Rotterdam, the Netherlands). At arrival in the laboratory, the FIT test was screened for collection date and presence of permission form and administrated in ICOLON IT database. ICOLON database was developed and owned by the regional organization for Population Screening South-West Netherlands, Rotterdam, the Netherlands. After arrival at the laboratory OC-sensor FIT test was stored at -20°C, FOB-gold FIT test was stored at 4 °C (median 3 days, range 1- 6 days) untill analysis. Both FITs were stored according to the manufacturers recommendations. FIT test, which were inappropriately used or non-analysable, were marked in ICOLON, and participants received a new FIT test. All other specimens received in the laboratory were analysed. If a specimen was received in the laboratory >7 days from date of sample collection and the test result was < 10 µg Hb/g feces, the participant received a new FIT test and the test result was discarded. If a specimen received in the laboratory >7 days from date of sample collection and test result was ≥ 10 µg Hb/g feces, the participant received a positive test result and a reference for colonoscopy. The number of and reasons for non-analysable FITs were recorded by the laboratory analysists.

The OC sensor FIT tests were analysed on two OC-sensor µ systems (Eiken, Japan), the FOB-gold FIT tests were analysed on a Sentinel Sentifit 270 system (Sentinel, Italy). All FIT tests were allowed to warm to room temperature before analysis and analysed once. The analytical working range was 1-200 µg Hb/g feces for the OC sensor µ, and 1-170 µg Hb/g feces for the Sentifit 270. Samples with Hb level above the upper analytical working limits were not diluted or re-analysed. Before the start of the study the OC sensor µ system and the Sentifit 270 system were compared. Fifty-five fecal samples were spiked with different concentrations of Hb, from each spiked sample two OC sensor Fit tests and two FOB-gold Fit tests were taken and analysis on OC sensor µ or Sentifit 270 respectively. Using paired Student T-test no significant difference was found by levels ≤ 65 µg Hb/g feces (p=0.412). The OC senor µ was calibrated with 6 calibrators. In the Erasmus MC this was done every week and in advance of every analytical run two quality controls (low and high) were measured. In the AMC the calibration took place when necessarily, indicated by the results of the

controls, and was monitored by two controls at the start and at the end of each run calibrators and controls were from Eiken, Japan. The Sentifit 270 was calibrated with 6 calibrators every month or earlier when another latex lot was used for analysis. Three quality controls (low, middle and high) were run every analytical run before and after analysis of the samples. Calibrators and controls were from Sentinel, Italy. Analyses were carried out in a fourteen-month period by seven technicians and two staff members, with over 8 years expertise in FIT analysis. Samples that appeared to be over range were not diluted and re-analysed.

Follow-up evaluation

Participants with a positive FIT result were scheduled for colonoscopy within 4 weeks. In case of an incomplete colonoscopy, a computed tomographic colonography was performed. Experienced endoscopists, all board-certified gastroenterologists who had performed at least 1,000 colonoscopies, performed all colonoscopies for the current trial. The maximum reach of the endoscope, quality of bowel preparation, data on location, size, macroscopic aspect, morphology, and endoscopic assessment of completeness of resection was recorded for all lesions detected during colonoscopy. All lesions were collected and evaluated by experienced gastrointestinal pathologists according to the Vienna criteria and World Health Organization classification[21,22]. Advanced adenomas (AA) were defined as an adenoma with a diameter ≥10 mm, and/or with a ≥25% villous component, and/or high-grade dysplasia. Advanced neoplasia (AN) included AA and CRC. Cancers were staged according to the 7th edition of the American Joint Committee on Cancer classification,[23]. Advice regarding surveillance colonoscopy after removal of adenomatous polyps, large (≥10 mm) serrated lesions or cancer was given to the clients according to the Dutch guideline. Participants with a negative colonoscopy were referred back to the screening program, but were considered not to require FIT-screening for 10 years.

Statistical analysis

The analysis was based on the intention-to-screen principle. The primary outcome measure was the diagnostic yield for advanced neoplasia, defined as the proportion of participants being diagnosed with AN relative to the total number of invitees. When more than one lesion was present, the screenee was classified according to the most advanced lesion.

Additional outcome measures were the participation rate, usability, the positivity rate, the positive predictive value for CRC and AN, the diagnostic yield of CRC defined as the proportion of participants with CRC relative to the number of invitees, and detection rate for AN and CRC, defined as the proportion of subjects with AN/CRC relative to all participants. The participation rate

was calculated as the number of participants relative to all eligible invitees. The PR was defined as the proportion of participants with a positive test result. Usability was defined as the number of non-analysable tests. The PPV refers to the participants in whom AN is detected relative to those undergoing colonoscopy after a positive FIT or, in case the colonoscopy was incomplete, computed tomographic colonography. As this is a dynamic cohort, outcomes were also separately analysed for first-time and repeat-round invitees. Adenoma detection rate was defined as the proportion of colonoscopies in which one or more adenomas were found.

Differences in proportions between groups were analysed for statistical significance using the χ2-test statistic. Differences in means between groups were tested using the Student's t statistic. Participation rate, PR, DR, and PPV were calculated and described as proportions with 95% confidence intervals (95% CI). All p values were two-sided and considered significant if <0.05. Analyses were conducted using SPSS for Windows version 21.0.

Sample size calculation

Sample size was guided by the size of the dynamic cohort, invited for the three previous screening rounds. This cohort consists of approximately 20,000 people. We anticipated that differences in participation would be crucial in driving any differences in diagnostic yield. Diagnostic yield would also be affected by failures, positivity, and positive predictive values. Inviting 20,000 people would then have a power of at least 80% to detect an absolute difference in diagnostic yield of 5 per 1,000 invitees or more, assuming 60% participation with FIT-based screening, a 6% positivity rate, and a 30% positive predictive value, using two-sided testing at a 5% significance level.

Ethical approval

The study was approved by the Dutch National Health Council (Population Screening Act; publication no. 2013/20).

Results

We invited 19,618 persons, of whom 327 had to be excluded because they met one of the exclusion criteria (n=306), had moved (n =19), or died (n=2), leaving 19,291 eligible invitees (Figure 1). Of the 9,669 invitees who received the OC-Sensor, 4,706 (49%) were male. Of the 9,622 invitees who received the FOB-Gold, 4,584 (48%) were male. The median age in both study arms was 61 years (IQR 57-67). The proportion of first-time invitees was 14.2% in the OC-sensor group and 14.8% in the

FOB-Gold group (p=0.23).

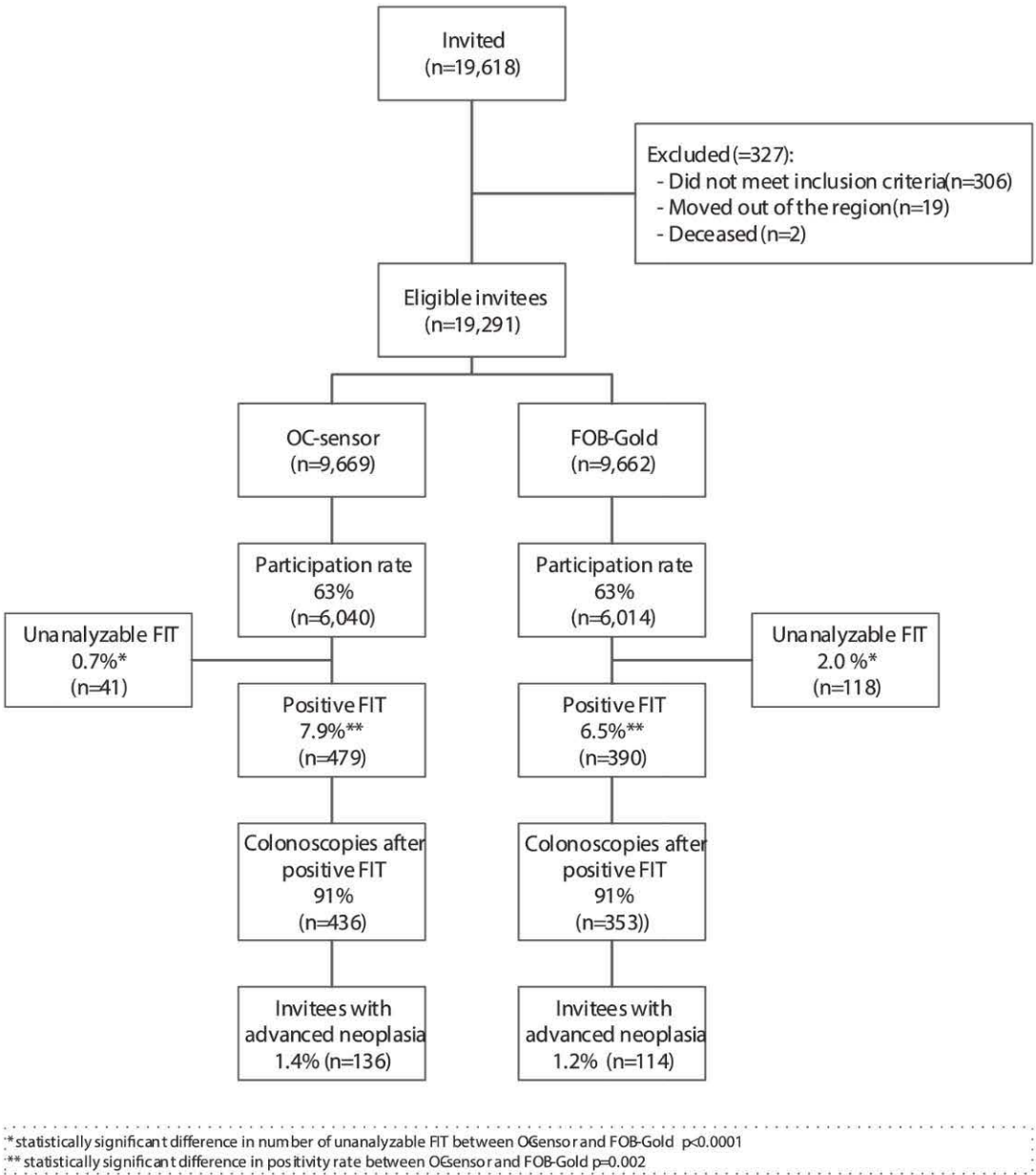


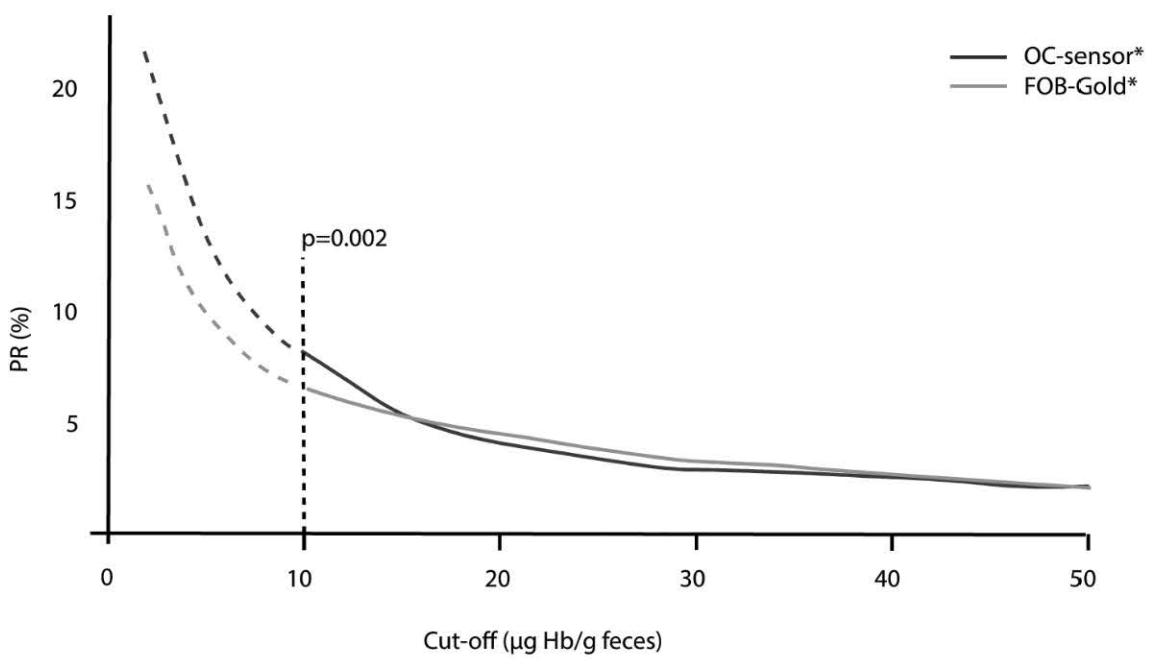
Figure 1. *Study design.*

In the OC-Sensor group, 6,040 returned the FIT (63%), versus 6,014 in the FOB-Gold group (63%) (p=0.96). Participation was lower among first-time invitees, yet over all rounds participation was similar for both FIT-brands (57% for OC-sensor and 56% for FOB-Gold; p=0.73). Non-analysable tests were reported in 41 (0.7%) participants in the OC-Sensor arm versus 118 (2.0%) in the FOB-Gold arm (p<0.001). The main reason for an unalyzable test was a too large sample of feces collected in the tube by the participant (Table1).

Table 1. *Reasons for failure to analyse FIT.*

Reasons	OC-sensor (%)	FOB-Gold (%)
	(n=41)	(n=118)
too large sample	40 (98)	97 (82)
too little sample	1 (2)	5 (4)
loss of buffer	0	7 (6)
analytical technical failure	0	9 (8)

More participants tested positive at the pre-specified cut-off of 10 µg Hb/g feces with the OC-Sensor: 479 (7.9%) versus 390 (6.5%) for FOB-Gold (p=0.002). The difference in positivity rate disappeared at higher cut-offs (Figure 2).



*dashed lines represent positivity rates of Hb concentrations under the pre-specified cut-off (10 µg Hb/gfeces)

Figure 2. *Positivity rates (PR) for the OC-sensor and FOB-Gold at different cut-offs.*

Positivity rates among first time invitees were higher and similar for both FITs, respectively 9.5% for OC-sensor and 9.1% for FOB-Gold (p=0.86). Fecal hemoglobin concentrations were distributed differently across both tests, with the OC-Sensor measuring higher values of Hb concentrations, ranging up to 1333 µg Hb/g feces, versus 179 µg Hb/g feces for FOB-Gold (p <0.0001, Figure 3).

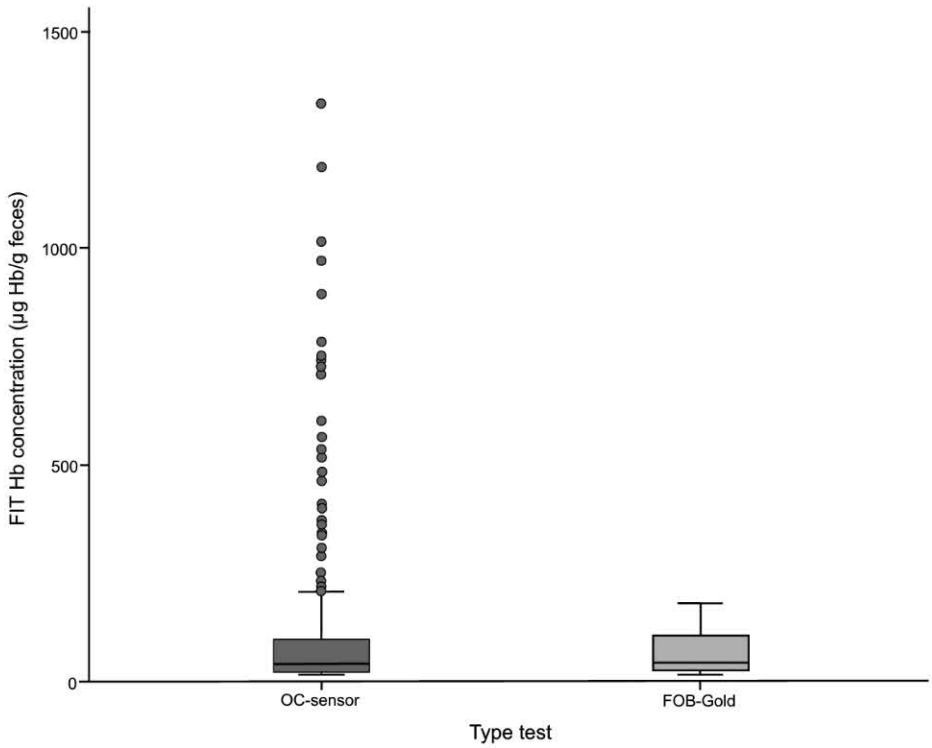


Figure 3. *Fecal hemoglobin concentrations of OC-sensor and FOB-Gold .*

Adherence to colonoscopy among FIT-positive screenees was 90% for both tests. Overall adenoma detection rate was 58%. Advanced neoplasia was detected in 137 participants in the OC-Sensor group (1.4%) and in 114 in the FOB-Gold group (1.2%; p=0.15). Specific colonoscopy findings are described in Table 2. In 537 FIT-positives (68%), no adenomas or only non-advanced adenomas were detected. Advanced adenomas were detected in 224 (28.4%) participants, and CRC was diagnosed in 27 (3.4%) participants, with most CRCs detected at an early stage (56% stage I). For the two FITs, no significant differences were observed in colonoscopy findings, including non-advanced adenomas, advanced adenomas and CRC, neither between stages of CRC.

Table 2*. Colorectal cancer stage for the OC-sensor and FOB-Gold among FIT-positive screenees.*

	OC-sensor (%)	**FOB-Gold (%)**	**p-value**
	(n=13)	(n=14)	
stage I	8 (61.5)	7 (50.0)	
stage II	1 (7.7)	1 (7.1)	0.63*
stage III	3 (23.1)	4 (28.7)	
stage IV	1 (7.7)	2 (14.3)	

* overall p-value

The diagnostic yield for CRC was 0.1% in both groups (p=0.84). The detection rate of AN among participants was 2.3% for OC-Sensor and 1.9% for FOB-Gold (p=0.15; Table 3).

Table 3. *Test characteristics for the OC-sensor and FOB-Gold for all invitees (n=19,291) and for first time invitees (n=2,796).*

	All invitees			**First-time invitees**		
	OC-sensor (n=9,669)	FOB-Gold (n=9,622)	p-value	OC-sensor (n=1,371)	FOB-Gold (n=1,424)	p-value
Positive predictive value (%)						
Adv. neoplasia	31.3	32.0	0.86	51.5	44.6	0.49
Colorectal cancer	3.0	4.0	0.45	6.1	6.2	1.0
Diagnostic yield (invitees, %)						
Adv. neoplasia	1.4	1.2	0.15	2.5	2.0	0.45
Colorectal cancer	0.1	0.1	0.84	0.3	0.3	1.0
Detection rate (participants, %)						
Adv. neoplasia	2.3	1.9	0.15	4.4	3.6	0.52
Colorectal cancer						

The detection rate for CRC was 0.2% for both tests (p=0.84). The PPV for AN was 31% for OC-Sensor and 32% for FOB-Gold at the pre-specified cut-off (p=0.80). For CRC the PPV was 3.0% for OC-Sensor versus 4.0% for FOB-Gold (p=0.45). In addition, as this study concerns a fourth round of screening we have stratified the results per number of participations to (previous) screening rounds for all participants of this fourth round (Table 4).

Both the PPV as well as the detection rate for AN was highest in those who participated for the first time. Because equal cut-offs did not result in equal PRs, we next calculated the PPV for different PRs using multiple cut-offs, this is illustrated in Figure 4. This yielded a similar PPV for both tests when comparing tests at the same PR's, and resulted in similar partial area's under the curve (p=0.48). This figure illustrates that both tests performed equally in terms of diagnostic yield for AN when comparing them at the same positivity rate and thus also at an equal number of colonoscopies required.

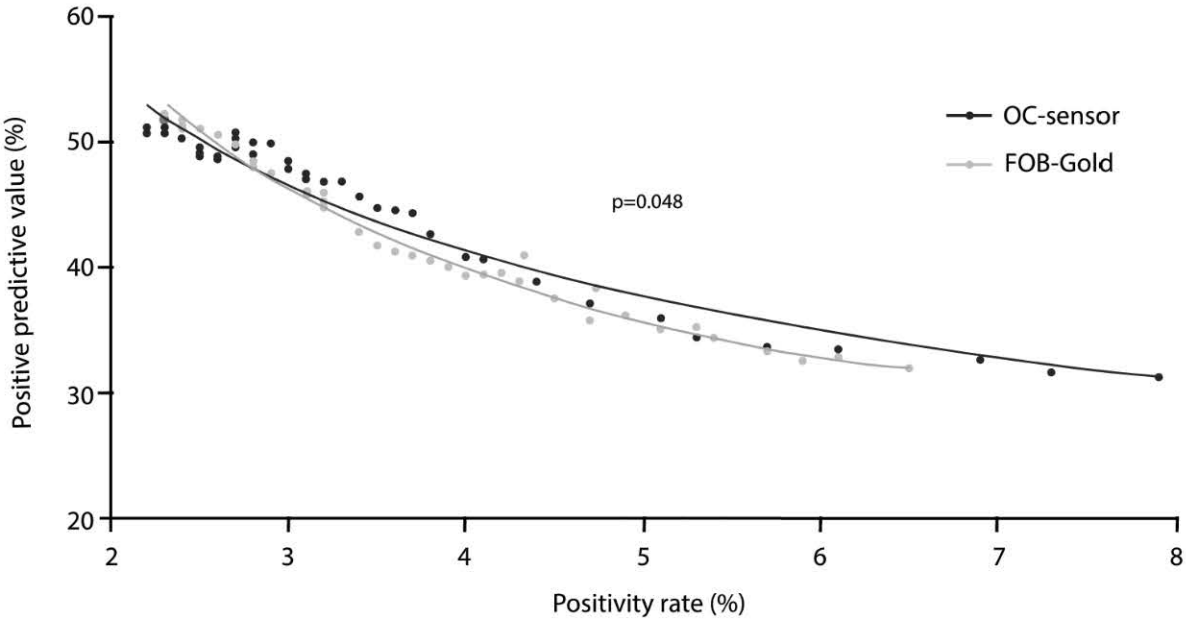


Figure 4. *Positive predictive value and positivity rate*

Discussion

In CRC screening programs, willingness to undergo a screening modality, easy use of the test and diagnostic accuracy are vital. In this cohort of biennial population-based FIT-screening, we observed similar performance of the OC-Sensor and FOB-Gold. Both tests had the same participation rate of 63%. A higher positivity rate was found for OC-Sensor, resulting in more colonoscopies and a slightly higher detection rate of AN. A significantly higher proportion of non-analysable tests was found with FOB-Gold. Both tests performed similar in terms of positive predictive value. However, screening with the OC-Sensor test led to a higher diagnostic yield, mainly due to the previously mentioned higher positivity rate. When comparing both tests at the same positivity rates by raising the cut-off of the OC-sensor, they showed similar positive predictive values.

Our study has several strenghts. All invitees were randomly selected from the general population and were at average risk of developing CRC. Invitees from the same household received the same test-brand, so that confusion about use or shape of the test was avoided. In addition, risk of exchanging the two FITs between participants living in the same household was thereby prevented. Adherence to colonoscopy after a positive FIT was high with 90% of screenees undergoing colonoscopy. A possible limitation for the results of our study is that it was performed in a 4th round of FIT-screening, with the consequence that the majority of the population was not screening-naïve. However, as

Table 4. *Positive predictive value and detection rate in participants according to number of participations over all rounds (n=12,054).*

	1st time participants (n=2,582)		**2nd time participants** (n=2,459)		**3rd time participants** (n=1,720)		**4th time participants** (n=5,293)	
	OC-sensor	FOB-Gold	OC-sensor	FOB-Gold	OC-sensor	FOB-Gold	OC-sensor	FOB-Gold
Positive predictive value (%)								
Advanced neoplasia	45.1	43.0	30.7	27.3	27.3	22.6	23.9	29.7
Colorectal cancer	5.7	4.7	0	7.6	-	-	3.3	3.1
Detection rate (participants, %)								
Advanced neoplasia	4.3	3.5	1.9	1.5	1.7	1.4	1.6	1.5
Colorectal cancer	0.5	0.4	0	0.4	-	-	0.2	0.2

the cohort included previous participants as well as newly invited individuals, it represents a true population-based screening population. Secondly, it is possible that participants were more familiar with the OC-sensor, which had been used in the first three rounds, and this may have influenced the usability of the FOB-Gold. However, a higher rate of non-analysable tests was also found among first time participants (1.0% for the OC-sensor vs. 2.9% for the FOB-Gold, p-value <0.001). Lastly, a more precise comparison of the diagnostic test accuracy of the two FITs could have been made by applying both tests on the same fecal samples, with all participants undergoing subsequent colonsocopy. This would have allowed to evaluate prime indicators of diagnostic performance, including sensitivity, specificity and the area under the receiver-operating curve. However, such a design might influence the willingsness to participate and would not have allowed for a fair comparison of participation rates and usability.

In our study, the main reason for non-analysable tests was a too large sample of feces collected by in the tube by the participant. These findings are in line with previously published results[24, 25]. A possible explanation for this sampling error could be the round shape of the opening of the FOB-Gold, making it possible to sample larger volumes of feces, whereas OC-Sensor has a small oval opening (Figure 5). It should be noted that the FOB-Gold test-tube that was used in our study has been adjusted before the start of the study by the manufacturer to prevent participants from unscrewing the wrong side of the test causing loss of buffer[26].

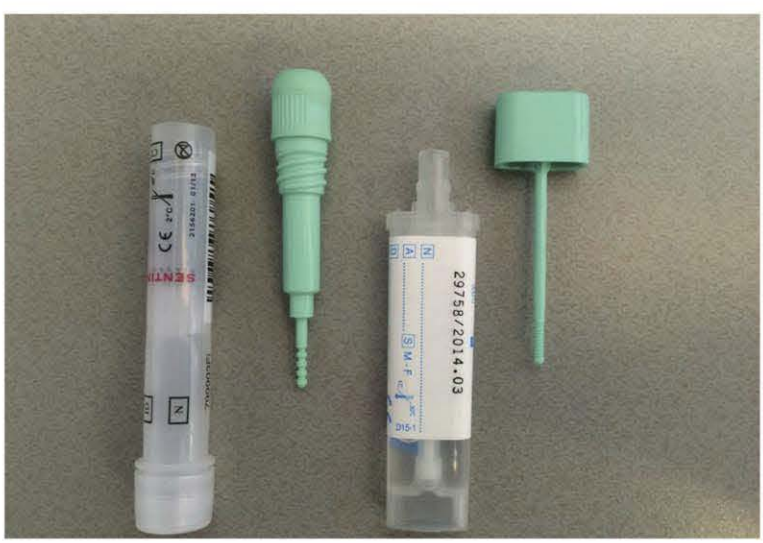


Figure 5. *The FOB-Gold (left) and OC-sensor (right).*

Concerns have been raised about the so-called prozone effect in FIT-screening[27, 28]. This effect could lead to an underestimation of high Hb concentrations because the relative large amount of antigen (in this case Hb) is greater than the quantity of antibody present in the test. This could lead to

underestimation of true Hb concentration at very high fecal Hb concentrations. Our results indicate that OC-Sensor is less subject to this effect than the FOB-Gold, with a large difference in distribution of the highest concentrations of Hb between both tests. The ability to measure high values is often driven by properties of the analytical equipment. However, it is important to realize that such high concentrations are usually much higher than most cut-offs, and are therefore not likely to influence positivity rates.

In our cohort OC-Sensor had a higher positivity rate than FOB-Gold at a cut-off of 10 µg Hb/g feces. This is in contrast with results from a Spanish study, in which the FOB-Gold had a higher positivity rate than the OC-Sensor[24]. This difference can be explained by multiple factors. First, in the Spanish study a higher positivity cut-off (100 ng Hb/ml) was used; we found that for higher cut-offs the difference in positivity rate between the tests was less pronounced. Second, the Spanish study used the same cut-off (expressed in ng Hb/ml buffer) for both tests. However after standardizing this to µg Hb/g feces, this cut-off relates for the OC-Sensor to a higher cut-off than for the FOB- Gold, 20 µg Hb/g and 17 µg Hb/g feces respectively[14, 29]. Last, the Spanish study was conducted in 2009, after which date both manufacturers have improved their buffer to increase the conservation of Hb in the sample.

Besides participation, test accuracy is crucial for screening effectiveness. In our study, the positive predictive value was comparable for both tests. In literature, the above mentioned Spanish study is the only other study that compared the same two brands of FIT in a screening population, allowing comparison of participation rates in addition to test accuracy. This study was performed in a first round of screening and demonstrated superiority of the OC-Sensor in participation rates, with similar positive predictive values for both tests[24]. Other studies that have evaluated the OC-sensor and/or FOB-Gold test relied on different designs to evaluate test performance. Such study designs however do not allow comparison of participation rates. One of those studies compared the two FITs to gFOBT, and concluded that apart from the superiority of FIT over gFOBT, no differences in performance were found between FITs[9]. These findings are in accordance with our results, confirming that similar test performance of both FITs can also be expected in a population-based screening setting. Another study sent both FITs to screenees, and in case of one or two positive test results participants were referred for colonoscopy. In this setting a relative sensitivity could be calculated, showing a lower sensitivity for the FOB-Gold, and comparable specificity,[30]. As all participants in our study only received one out of two tests, sensitivity could not be calculated. However, as both groups were randomized, prevalence of AN should be similar between groups and

as a result the detection rate is a direct reflection of the relative sensitivity. Also, future follow-up will allow determining the incidence of interval carcinomas to determine program sensitivity per test.

As it has become clear that different FITs use a variety of sampling techniques and report Hb concentration in different units, more attention has been given to standardizing fecal hemoglobin concentrations[14]. Due to these differences a comparison between FITs could prove to be arduous. It has been proposed to standardize the measuring units to µg Hb/g feces, taking into account the amount of buffer and sampling volume[14]. However, a Taiwanese study concluded that even after standardization, different brands of quantitative FITs perform differently[13]. Our findings are in line with these results, showing different positivity rates and PPVs per test at the same cut-off. This is likely explained by the fact that standardizing FITs to µg Hb/g feces is hampered by several factors. Firstly, wet sample FITs do not directly determine hemoglobin concentration in feces, but determine hemoglobin concentration in the kit's storage buffer. This depends on both the fecal hemoglobin concentration and the amount of fecal material put into the buffer. Although manufacturers assume that the volume of fecal material sampled is stable per device, sampling volumes can in practice vary substantially. This affects the reported fecal hemoglobin concentrations. Secondly, different FIT brands make use of antibodies against different epitopes. This could potentially influence test performance and positivity rate. As a result, the same cut-off in µg Hb/g feces can lead to different positivity rates depending on the FIT brand. Therefore we chose to compare the positivity rate between both FITs, this resulted in comparable PPVs. Another advantage of this approach is that the positivity rate directly reflects the required colonoscopy capacity. Evidently, a higher positivity rate leads to more colonoscopies and consequently results into a higher detection rate of AN in case of similar PPV's. As colonoscopy capacity is for most countries the main determinant in a CRC screening program, information on required capacity using test positivity rate is crucial when implementing a population-based screening program or when contemplating on changing to a different FIT brand within an already existing program[12]. This necessitates a comparative analysis of test performances not solely based on similar cut-offs but firstly based on similar PRs. Moreover, if programs intent to use different cut-off concentrations for different populations (e.g., based on age, gender or screening history) PRs for the different cut-offs per subgroup should be addressed. Our comparison shows that both tests perform equally regarding detection rate and PPV at cut-offs that result in equal PRs requiring the same number of colonoscopies. Notably, miss rates of early stage cancers between tests is another important outcome, the fact that both tests led to very similar numbers of cancers detected strongly suggests that they will also perform similar in terms of interval cancers. This trial shows that the OC-Sensor and FOB-Gold can be expected to perform similar in population-

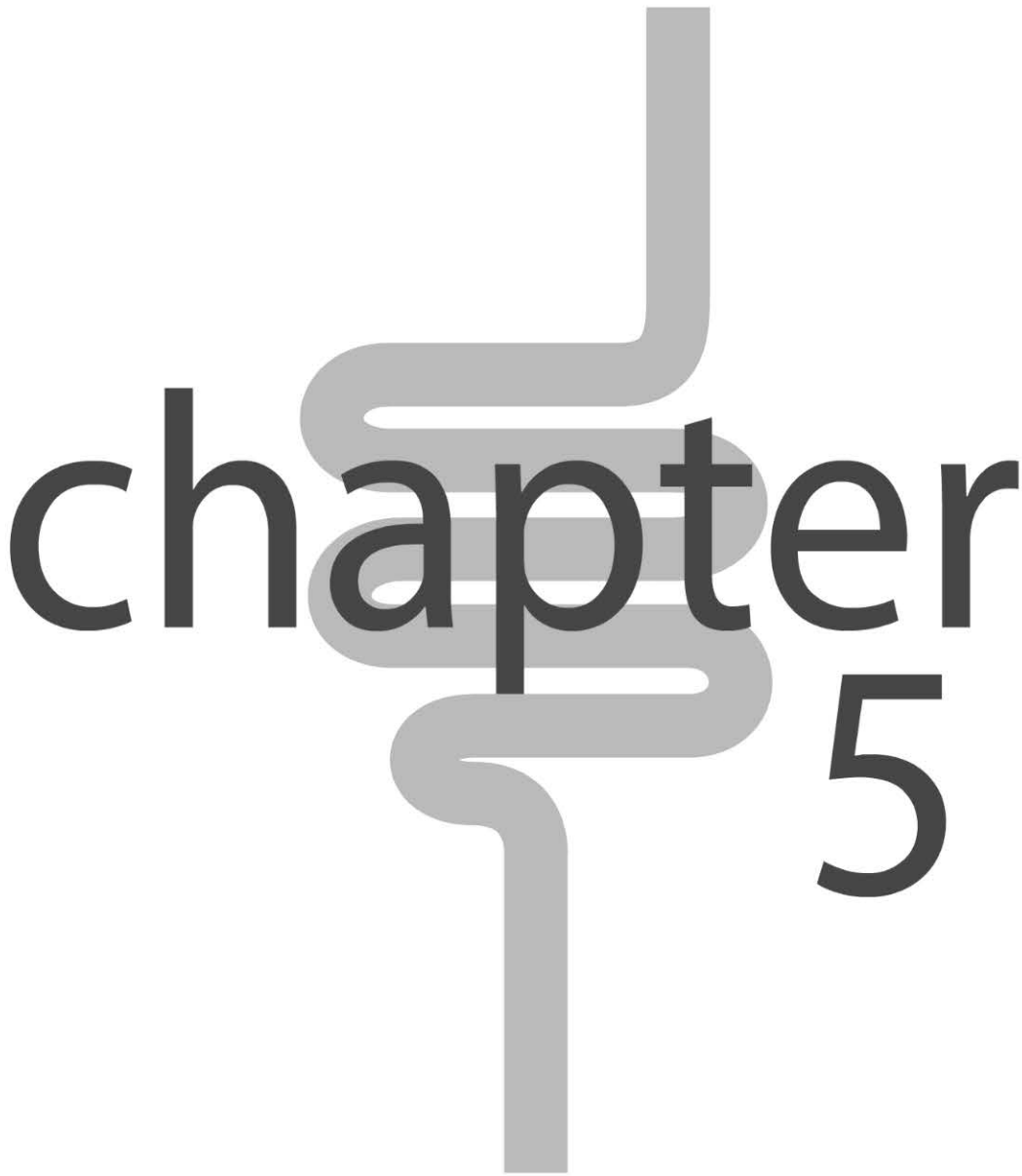
based screening, with no major differences in diagnostic yield. Despite standardization on Hb concentration, differences in positivity rate and diagnostic yield can be expected, but adjusting for PR will result in an equal number of colonoscopies, and a similar diagnostic yield. When comparing different FIT-brands, our results indicate a need for standardizing on positivity rate, rather than on fecal hemoglobin concentration.

References

1. Center MM, Jemal A, Ward E. International trends in colorectal cancer incidence rates. Cancer Epidemiol Biomarkers Prev 2009; 18(6): 1688-94.
2. Edwards BK, Ward E, Kohler BA, et al. Annual report to the nation on the status of cancer, 1975-2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. Cancer 2010; 116(3): 544-73.
3. Kuipers EJ, Grady WM, Lieberman D, et al. Colorectal cancer. Nature Reviews Disease Primers 2015: 15065.
4. Shaukat A, Mongin SJ, Geisser MS, et al. Long-term mortality after screening for colorectal cancer. N Engl J Med 2013; 369(12): 1106-14.
5. Hol L, de Jonge V, van Leerdam ME, et al. Screening for colorectal cancer: comparison of perceived test burden of guaiac-based faecal occult blood test, faecal immunochemical test and flexible sigmoidoscopy. Eur J Cancer 2010; 46(11): 2059-66.
6. Brenner H, Tao S. Superior diagnostic performance of faecal immunochemical tests for haemoglobin in a head-to-head comparison with guaiac based faecal occult blood test among 2235 participants of screening colonoscopy. Eur J Cancer 2013.
7. van Rossum LG, van Rijn AF, Laheij RJ, et al. Random comparison of guaiac and immunochemical fecal occult blood tests for colorectal cancer in a screening population. Gastroenterology 2008; 135(1): 82-90.
8. Rabeneck L, Rumble RB, Thompson F, et al. Fecal immunochemical tests compared with guaiac fecal occult blood tests for population-based colorectal cancer screening. Can J Gastroenterol 2012; 26(3): 131-47.
9. Faivre J, Dancourt V, Denis B, et al. Comparison between a guaiac and three immunochemical faecal occult blood tests in screening for colorectal cancer. European Journal of Cancer 2012; 48(16): 2969-76.
10. Hol L, van Leerdam ME, van Ballegooijen M, et al. Screening for colorectal cancer: randomised trial comparing guaiac-based and immunochemical faecal occult blood testing and flexible sigmoidoscopy. Gut 2010; 59(1): 62-8.
11. 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.
12. Schreuders EH, Ruco A, Rabeneck L, et al. Colorectal cancer screening: a global overview of existing programmes. Gut 2015; 64(10): 1637-49.
13. Chiang TH, Chuang SL, Chen SL, 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.
14. Fraser CG, Allison JE, Halloran SP, Young GP, Org WE. A Proposal to Standardize Reporting Units for Fecal Immunochemical Tests for Hemoglobin. Jnci-J Natl Cancer I 2012; 104(11): 810-4.
15. Lee JK, Liles EG, Bent S, Levin TR, Corley DA. Accuracy of fecal immunochemical tests for colorectal cancer: systematic review and meta-analysis. Ann Intern Med 2014; 160(3): 171.
16. Stegeman I, de Wijkerslooth TR, Mallant-Hent RC, et al. Implementation of population screening for colorectal cancer by repeated Fecal Immunochemical Test (FIT): third round. BMC Gastroenterol 2012; 12: 73.
17. Kapidzic A, Grobbee EJ, Hol L, et al. Attendance and yield over three rounds of population-based fecal immunochemical test screening. Am J Gastroenterol 2014; 109(8): 1257-64.
18. Denters MJ, Deutekom M, Fockens P, Bossuyt PM, Dekker E. Implementation of population screening for colorectal cancer by repeated fecal occult blood test in the Netherlands. BMC Gastroenterol 2009; 9: 28.
19. Denters MJ, Deutekom M, Bossuyt PM, Stroobants AK, Fockens P, Dekker E. Lower risk of advanced neoplasia among

patients with a previous negative result from a fecal test for colorectal cancer. Gastroenterology 2012; 142(3): 497-504.

20. Stegeman I, van Doorn SC, Mundt MW, et al. Participation, yield, and interval carcinomas in three rounds of biennial FIT-based colorectal cancer screening. Cancer Epidemiol 2015; 39(3): 388-93.
21. Schlemper RJ, Riddell RH, Kato Y, et al. The Vienna classification of gastrointestinal epithelial neoplasia. Gut 2000; 47(2): 251-5.
22. Hamilton SR, Rubio CA, B. V, et al. World Health Organization classification of tumours. Tumours of the digestive system. Lyon, France, 2000.
23. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. Ann Surg Oncol 2010; 17(6): 1471-4.
24. Zubero MB, Arana-Arri E, Pijoan JI, et al. Population-based colorectal cancer screening: comparison of two fecal occult blood test. Front Pharmacol 2014; 4: 175.
25. Santare D, Kojalo I, Huttunen T, et al. Improving uptake of screening for colorectal cancer: a study on invitation strategies and different test kit use. Eur J Gastroenterol Hepatol 2015; 27(5): 536-43.
26. FOB-Gold, Hb Fecal Immunochemical Test - Sentinel Diagnostics. http://www.sentinel.it/en/fob-gold-line/fobgold.aspx (accessed July 14 2015).
27. Vaananen P, Tenhunen R. Rapid immunochemical detection of fecal occult blood by use of a latex-agglutination test. Clin Chem 1988; 34(9): 1763-6.
28. Carroll M, Piggot C, Pearson S, Seaman H, Haloran SP. Evaluation of quantitative faecal immunochemical tests for haemoglobin: NHS Bowel Cancer Screening Southern Programme Hub, 2014.
29. Guittet L, Guillaume E, Levillain R, et al. Analytical Comparison of Three Quantitative Immunochemical Fecal Occult Blood Tests for Colorectal Cancer Screening. Cancer Epidem Biomar 2011; 20(7): 1492-501.
30. Rubeca T, Rapi S, Confortini M, et al. Evaluation of diagnostic accuracy of screening by fecal occult blood testing (FOBT). Comparison of FOB Gold and OC Sensor assays in a consecutive prospective screening series. Int J Biol Markers 2006; 21(3): 157-61.


chapter
5

Attendance and yield over three rounds of population-based fecal immunochemical test screening

Atija Kapidzic, Esmée J. Grobbee, Lieke Hol, Aafke H.C. van Roon, Anneke J. van Vuuren, Wolfert Spijker, Kirsten Izelaar, Marjolein van Ballegooijen, Ernst J. Kuipers, and Monique E. van Leerdam

American Journal of Gastroenterology, 2014

Abstract

Introduction Fecal immunochemical test (FIT) screening for colorectal cancer (CRC) requires timely successive rounds for an optimal preventive effect. However, data on attendance and trend in yield over multiple rounds of FIT screening are limited. We therefore conducted a consecutive third round of FIT screening in a population-based CRC screening trial.

Methods Average-risk subjects aged 50-74 years were approached for three rounds of 1-sample FIT (OC-sensor, Eiken Chemical, Japan) screening. Subjects with a hemoglobin level ≥ 50 ng/ml (≥ 10 µg hemoglobin/g feces) were referred for colonoscopy. Subjects with a positive FIT in previous rounds were not re-invited for FIT screening.

Results In the first round 7,501 subjects were invited. Participation rate was 62.6% in the first, 63.2% in the second and 68.3% in the third round ($p<0.001$). In total, 73% (5,241/7,229) of all eligible subjects participated in at least one of three rounds. The positivity rate was significantly higher in the first (8.4%) compared to the second (6.0%) and third (5.7%) screening round ($p<0.001$). The detection rate of advanced neoplasia declined from the first to subsequent rounds (round 1: 3.3%; round 2: 1.9%; round 3: 1.3%, $p<0.001$). The positive predictive value for advanced neoplasia was 40.7% in the first, 33.2% in the second and 24.0% in the third screening round ($p<0.001$).

Conclusion Repeated biennial FIT screening is acceptable with increased participation in successive screening rounds, and more than 70% of all eligible subjects participating at least once over three rounds. The decline in screen-detected advanced neoplasia over three screening rounds is compatible with a decreased prevalence of advanced neoplasia as a result of repeated FIT screening. These findings provide strong evidence for the effectiveness of FIT screening and stress the importance of on-going research over multiple screening rounds.

Introduction

Colorectal cancer (CRC) is a major health concern worldwide. Screening using fecal occult blood testing (FOBT) results in detection and treatment of CRC at an earlier stage, which is associated with improved survival. A meta-analysis based on four large randomized controlled trials demonstrated that guaiac fecal occult blood (gFOBT) testing reduces CRC-related mortality[1]. A recently published study showed that the effect of gFOBT screening on CRC mortality persists for many years[2]. Economic analyses found gFOBT screening to be cost-effective[3]. More recently, fecal immunochemical testing (FIT) gained ground based on randomized trials showing higher attendance as well as a higher sensitivity for detection of advanced neoplasia with a similar specificity[4,5]. Moreover, quantitative measurement of fecal human globin concentrations offers the opportunity to provide tailored screening for specific regions or countries based on available colonoscopy capacity and cost-effectiveness analyses[3,6]. FIT screening has therefore become the first-choice fecal occult blood test for CRC screening[7]. Various CRC screening programmes worldwide currently rely on FIT or are about to start with or switch to FIT-based screening.

The sensitivity of a single round of FIT screening for the detection of advanced neoplasia is however limited. Recent studies showed that FIT at a low cut-off detects approximately 85% of CRCs and up to 35% of large adenomas[8,9]. Successive screening rounds are required to optimize the impact of FIT screening on a population level. Participation and detection rates in successive rounds attribute to the effectiveness of FIT-based programmes. Longitudinal adherence of the same subjects represents a critical factor, but information concerning sustained attendance and diagnostic performance over repeated rounds of FIT screening is very limited.

We and others demonstrated a stable attendance rate over two rounds of FIT screening, with detection of substantial numbers of advanced lesions in both rounds[10,11]. Data on further rounds in FIT screening with longer follow-up periods are scarce. One relatively small Italian study on four rounds of a biennial FIT screening programme reported stable attendance rates and test performances. However, the attendance rate of 56% during the first round was relatively low[12]. Further data on repeated FIT screening are warranted, as these provide more insight in the programme sensitivity of FIT screening. Such information is also required to address the important question whether FIT screening with higher sensitivity for advanced neoplasia can be applied with longer screening intervals than biennial gFOBT screening[7].

We therefore evaluated attendance and detection rates of three rounds of FIT screening in a Dutch population-based CRC screening programme.

Methods

Study population and study design

Details about the design of this on-going population-based CRC screening programme have been described previously 4,10. In short, demographic data of all individuals between 50-74 years living in the southwest of the Netherlands were obtained from municipal population registers. Random samples were taken from the target population by a computer-generated algorithm (Tenalea, Amsterdam, The Netherlands). Since there was no CRC screening programme at the time of the trial in the Netherlands, the target population was screening-naïve when first approached. Individuals with a history of inflammatory bowel disease or CRC, as well as those who had undergone a colonoscopy, sigmoidoscopy or barium contrast enema in the last 3 years, those with an estimated life expectancy of less than 5 years, and subjects who were unable to give informed consent were excluded from the study. Subjects were not invited for the third screening round in case of a positive FIT in the first or second screening round, when they had become older than 74 years, when they had moved out of the region, or when they had died. Recruitment took place between November 2006 and December 2012.

Intervention; FIT screening

With each screening round, one FIT (OC-sensor, Eiken Chemical, Japan) was sent by mail to collect a single sample of one bowel movement. The test result was considered positive when the hemoglobin concentration in the FIT sample was ≥ 50 ng/ml, which corresponds to ≥ 10 µg hemoglobin/g feces. Study subjects were initially divided over three groups to undergo repeated FIT testing at different screening intervals in the second round (ie, one, two and three years, respectively). No differences in attendance and detection rate were found between the different intervals 10. The positive predictive value did also not differ between the three screening intervals (one-year interval: 36.2%; two-year interval: 32.9%; three-year interval: 30.6%; p=0.773) (derived from 10). We therefore included subjects with a one-, two-, and three-year interval between the first and second round. Based on these results, a two-year interval was applied to all groups in the third screening round. In total, 5,482 subjects were invited for third round screening (1,838 subjects in the group with a previous one-year interval between the first and second round; 1,835 subjects with a two-year interval; and 1,809 subjects with a three-year interval).

Follow-up evaluation; colonoscopy

Subjects with a positive FIT were scheduled for colonoscopy within 4 weeks. In case the colonoscopy

was incomplete a CT-colonoscopy was performed. Experienced endoscopists, all board-certified gastroenterologists who had performed at least over 1,000 colonoscopies, performed all colonoscopies for the current trial. The maximum reach of the endoscope, adequacy of bowel preparation as well as the characteristics and location of any polyps were recorded. Gastrointestinal pathologists evaluated all removed polyps. Patients with a positive colonoscopy entered a surveillance programme according to guidelines of the Dutch Society of Gastroenterology, while subjects with a negative colonoscopy were referred back to the screening programme, but were considered not to require FIT screening for ten years.

Screen-detected and interval carcinomas

Except for individuals who had moved out of the Netherlands, all recruited participants were followed for the development of CRC. Screen-detected cancers were defined as cancers identified at colonoscopy performed after a positive test result. Interval cancers were defined as colorectal cancers diagnosed within the time period between attendances to screening. Interval cancers were identified through linkage with the Dutch Comprehensive Cancer Centre (www.iknl.nl).

Statistical analysis

For each screening round, we calculated the attendance rate (AR), the positivity rate (PR), the detection rate (DR) of CRC and advanced adenomas, and the positive predictive value (PPV) for CRC and advanced adenomas.

The AR was calculated by dividing the number of participants by all eligible subjects (defined as all invitees minus the excluded subjects). The PR was defined as the proportion of participants having a positive test result. The DR was defined as the proportion of participants being diagnosed with advanced neoplasia. This was calculated as the number of screened individuals diagnosed with advanced neoplasia divided by all screened individuals with an analyzable FIT test. Advanced neoplasia included CRC and advanced adenomas. An advanced adenoma (AA) was defined as an adenoma with a diameter ≥ 10 mm, and/or with a ≥ 25% villous component, and/or high grade dysplasia. When more than one lesion was present, the screenee was classified according to the most advanced lesion. The PPV refers to the subjects diagnosed with advanced neoplasia (AN) proportionally to screenees with a positive FIT undergoing subsequent colonoscopy, or in case the colonoscopy was incomplete a CT-colonoscopy (n=3). For the overall adenoma detection rate (ADR), we used both advanced and non-advanced adenoma.

Differences in proportions between groups were analyzed by Chi-square testing. Differences in means between groups were tested using the Student t-test. AR, PR, DR, and PPV were calculated

and described as proportions with 95% confidence intervals (95% CI). We fitted a logistic regression model to the data to determine differences in participation and FIT characteristics between the different groups that attended the third screening round (ie, subjects that had participated one, two or three times over the three rounds). To determine the number of true positives per 1,000 invitees (subjects with a positive FIT identified with AN during follow-up colonoscopy) per screening round, the participation rate was multiplied by the positivity rate and the positive predictive value. The percentage of stable attenders was defined as the number of subjects attending all rounds while they were eligible, divided by the total amount of subjects that were eligible over the three rounds. The cumulative attendance was defined as the number of eligible invitees attending at least once. To assess differences in attendance rate between the three rounds, a generalized estimating equation was used to account for clustering at the level of the invitee. The test characteristics in the first two rounds of 1-sample FIT screening were compared to those in the third screening round by using a logistic regression model 10. The diagnostic yield was compared to that of different CRC screening methods. All p-values were two-sided and considered significant if < 0.05. All tests were conducted using SPSS version 20.0.

Ethical approval

The Dutch National Health Council and the Institutional Review Board of the Erasmus MC University Medical Centre approved the study. All screenees gave written informed consent.

Results

Attendance

Baseline characteristics and the results of the first and second 1-sample FIT screening rounds have previously been described 4,10. Briefly, during the first round, a total of 7,501 average-risk subjects were invited to participate in screening. Participation rates in the first, second and third round were 62.6% (4,523/7,229, 95% CI 61.4-63.7), 63.2% (3,864/6,111, 95% CI 62.0-64.4), and 68.3% (3,704/5,423, 95% CI 67.1-69.5), respectively ($p<0.001$) (Table 1). Figure 1 shows the trial profile for each screening round for the three groups (group I: one-year interval between the first and second round; group II: two-year interval between the first and second round; group III: three-year interval between the first and second round; group I-III: two-year intervals between the second and third round). Seventy-three percent (5,482/7,501) of the initial cohort was eligible to be invited for the third screening round. In total, 1,247 subjects were not eligible for successive screening rounds because they had become 75 years or older (round 2: n=342; round 3: 295), or had had a positive

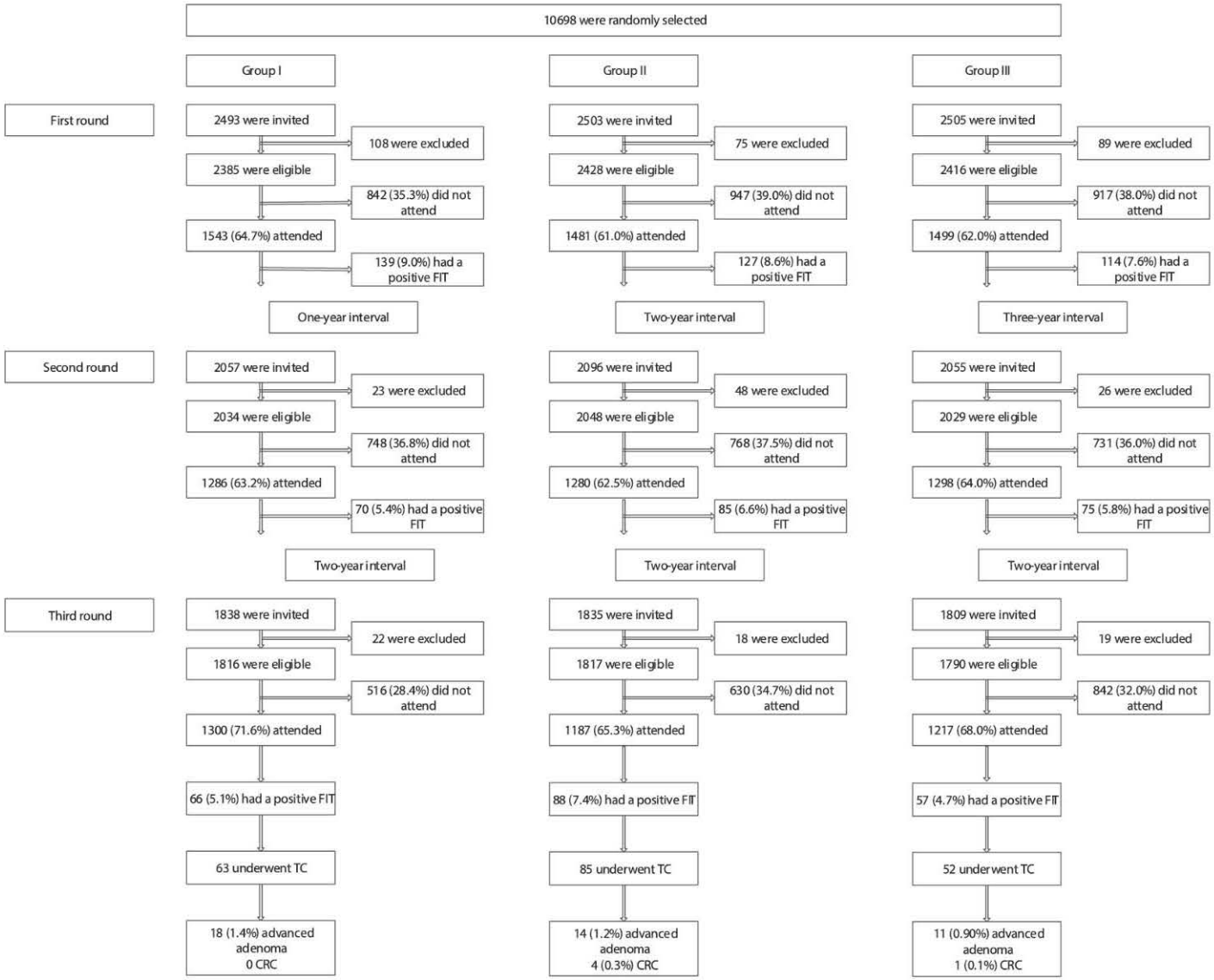


Figure 1. *Flowchart of study overview.*

FIT in previous rounds (round 1: n=380; round 2: n= 230). In addition, subjects were excluded during the first or second round because they had moved away (n=233), had died (n=170), or met one of the exclusion criteria (n=369).

In total, 5,482 subjects were invited to attend the third screening round (Table 1). A total of 59 subjects (1.1%) were excluded (47 subjects met one of the exclusion criteria, eleven had moved away and one had died) (Figure 1). Out of 5,423 eligible invitees, 3,704/5,423 (68.3%; 95% CI 67.1-69.5) returned a FIT. The test was analyzable in 3,700 (99.9%) subjects. Of the participants in the third round, 78.5% (2,907/3,704) attended all three rounds, 13.9% (516/3,704) attended two rounds, and 7.6% (281/3,704) had attended no previous round (Table 3). With respect to the non-participants in first-round screening, 18.8% (437/2330, 95% CI 17.2-20.4) attended the second round, while 23.2% (471/2,031, 95% CI 21.4-25.1) of second round non-participants attended the third round.

Table 1. *Overview of participation and FIT performance characteristics per screening round*

	Round 1	**Round 2**	**Round 3**	**p-value**
	% (95% CI)	% (95% CI)	% (95% CI)	
	(n=7229)	(n=6111)	(n=5423)	
Participation rate	62.6 (61.4-63.7)	63.2 (62.0-64.4)	68.3 (67.1-69.5)	<0.001
Positivity rate	8.4 (7.6-9.2)	6.0 (5.2-6.7)	5.7 (5.0-6.5)	<0.001
Colonoscopies (n)	364 (95.8%)	223 (97.0%)	200 (94.8%)	
Detection rate				
Non-advanced neoplasia	1.7 (1.3-2.1)	1.2 (0.9-1.6)	1.6 (1.2-2.1)	0.259
Advanced neoplasia	3.3 (2.8-3.8)	1.9 (1.5-2.4)	1.3 (1.0-1.7)	<0.001
Advanced adenoma	2.8 (2.3-3.3)	1.7 (1.4-2.2)	1.2 (0.9-1.6)	<0.001
Colorectal cancer	0.5 (0.3-0.7)	0.2 (0.1-0.4)	0.1 (0.1-0.3)	0.007
Positive predictive value				
Advanced neoplasia	40.7 (35.7-45.8)	33.2 (27.3-39.6)	24.0 (18.6-30.4)	<0.001
Advanced adenoma	34.6 (29.9-39.7)	30.0 (24.4-36.4)	21.5 (16.4-27.7)	0.005
Colorectal cancer	6.0 (4.0-9.0)	3.1 (1.5-6.4)	2.5 (1.0-5.9)	0.094

The number of times invitees participated during three screening rounds displayed for the number of times invitees were eligible is summarized in Table 2. In total 7,229 of 7,501 invitees were at least once eligible for screening. The proportion of stable attenders (ie, invitees attending all rounds while they were eligible) was 56.3% (4,073/7,229) (Table 2). The cumulative attendance rate (ie, eligible invitees attending at least one screening round) was 72.5% (5,241/7,229).

Table 2. *Number of times invitees participated displayed for the number of times they were eligible*

Times eligible	Times participated				Total
	0	1	2	3	
0	272 (100)	-	-	-	272
1	374 (33.6)	740 (66.4)	-	-	1114
2	181 (26.2)	85 (12.3)	426 (61.6)	-	692
3	1433 (26.4)	473 (8.7)	610 (11.2)	2907 (53.6)	5423
Total	**2260**	**1298**	**1036**	**2907**	**7501**

Proportion of positive tests

In total, 380/4,523 (8.4%, 95% CI 7.6-9.2%) tested positive in the first round, 230/3,864 (6.0%, 55% CI 5.2-6.7) in the second round, and 211/3,704 (5.7%, 95% 5.0-6.5) in the third round (Table 1). The positivity rate (PR) was significantly higher in the first compared to the second and third round (both p<0.001), whereas the PR was similar in round two and three (p=0.67). Individuals that participated for the first time in the third round had a significantly higher PR compared to individuals who underwent repeated screening (p<0.001) (Table 3).

Table 3. *Participation and FIT performance characteristics for subjects who have been invited in all three screening rounds (n=5482), by number of consecutive tests performed (ie, one, two or three out of three rounds)*

	Times participated*			p-value
	1 out of 3 rounds	2 out of 3 rounds	3 out of 3 rounds	
	(n=281)	(n=516)	(n=2907)	
Positive tests in the third round (n)	33	45	133	
% (95% CI)	11.7 (8.5-16.1)	8.7 (6.6-11.5)	4.6 (3.9-5.4)	<0.001
Colonoscopies performed n (%)	31 (94.0)	43 (95.6)	126 (94.7)	
Detection rate				
Advanced neoplasia (n)	13	7	28	
% (95% CI)	4.6 (2.7-7.8)	1.4 (0.6-2.8)	1.0 (0.7-1.4)	<0.001
Advanced adenoma (n)	12	6	25	
% (95% CI)	4.3 (2.4-7.4)	1.2 (0.5-2.6)	0.9 (0.6-1.3)	<0.001
Colorectal cancer (n)	1	1	3	
% (95% CI)	0.4 (0.1-2.5)	0.2 (0.0-1.4)	0.1 (0.0-0.3)	0.32
Positive predictive value				
Advanced neoplasia % (95% CI)	41.9 (26.1-59.6)	16.3 (8.0-30.4)	22.2 (15.8-30.3)	0.01
Advanced adenoma % (95% CI)	38.7 (23.5-56.5)	14.0 (6.4-27.8)	19.8 (13.8-27.7)	0.01
Colorectal cancer % (95% CI)	3.2 (0.5-19.6)	2.3 (0.3-14.7)	2.4 (0.8-7.1)	

* only participants eligible for three rounds

Follow-up and test performance characteristics

The detection rate (DR) and positive predictive value (PPV) for advanced neoplasia for the three rounds are described in Table 1. In the third round, 200 (94.8%) of 211 screenees that tested positive underwent a complete colonoscopy. The remaining 11 subjects either refused colonoscopy (n=10), or turned out to have too severe co-morbidity to benefit from an endoscopic procedure (n=1). The DR of advanced neoplasia (AN) was 3.3% (95% CI 2.8-3.8) in the first, 1.9% (95% CI 1.5-2.4) in the second, and 1.3% (95% CI 1.0-1.7) in the third round (p<0.001). The PPV for AN was 40.7% (95% CI 35.7-45.8) in the first, 33.2% (95% CI 27.3-39.6) in the second, and 24.0% (95% CI 18.6-30.4) in the third round (p<0.001) (Table 1). The DR declined significantly over the three screening rounds. In addition, the PPV only differed significantly between the second and third screening round (p=0.02), but not between first and second round screening (p=0.07). The overall adenoma detection rate (ie, of both advanced and non-advanced adenoma) in this study over three screening rounds was 57.4% (95% CI 53.9-60.8).

Both the DR and PPV were significantly higher in individuals that participated for the first time (Table 3, subgroup 1 out of 3 rounds) compared to individuals that underwent repeated screening (Table 3: subgroup 2 out of 3 rounds and subgroup 3 out of 3 rounds; DR: p<0.001; PPV: p=0.01).

The number of true positives (subjects with a positive FIT identified with AN during follow-up colonoscopy) per 1,000 subjects invited was 21 in the first round and 34 after two consecutive screening rounds 7. After three consecutive rounds of FIT screening, this number was 43 per 1,000 invitees.

Interval carcinomas

After record linkage with the Dutch Comprehensive Cancer Centre, 43 CRCs were found in the total study population. Thirty-four CRCs (79.1%) were screen-detected tumors, of which 22 (65%) were detected in the first, seven (21%) were detected in the second and five (15%) were detected in the third screening round. The remaining nine (20.9%) were interval carcinomas of which three were detected between the first and second round and six between the second and third round. Two of the interval cancers between the first and second round were detected in participants with a negative FIT: one Stage III tumor (FIT result at baseline, 24 ng/ml) was detected nine months after baseline screening, and one stage II tumor (FIT result at baseline, 7 ng/ml) was diagnosed two years and five months after FIT screening. The third CRC was diagnosed at stage I in a subject with a positive FIT but negative colonoscopy 10. One year and four months after the index colonoscopy, a subsequent colonoscopy was performed because of symptoms and revealed a tumor located at 50

cm of the anal verge. Two of the interval cancers between the second and third round were detected in participants with a negative 2nd round FIT: one stage IV tumor was detected 12 months after 2nd round screening (FIT result at 2nd round, 48 ng/ml), and one stage III tumor was detected 5 months after 2nd round FIT screening (FIT result at 2nd round, 0 ng/ml). Two of the interval cancers were diagnosed in subjects with a positive 2nd round FIT but negative subsequent colonoscopy. In one subject a second colonoscopy because of symptoms revealed a stage III CRC in the sigmoid twenty-four months after the index colonoscopy. A stage I CRC was diagnosed in the splenic flexure thirty-six months after the index colonoscopy. The fifth and sixth CRC were diagnosed in subjects who had a negative FIT result in the first round (24 respectively 0 ng/ml), were ineligible for the second round due to age, and developed a CRC two years and ten months (stage II) respectively three years and two months (stage III) after first round screening.

Discussion

This is a population-based study on the performance of repeated FIT screening with three rounds. Given the scarcity of information on impact of repeated FIT screening, such data are of major importance for countries considering or planning the implementation of population-based FIT screening. We observed a high and increasing attendance and a decline in detection rate of advanced adenoma over three consecutive rounds.

A very important early indicator for an effective population-based screening programme is uptake. We observed a high attendance per screening round that increased over successive screening rounds. Uptake in previous FOBT studies varied, but data of our group and others showed through randomized studies that FIT screening results in a higher uptake compared to gFOBT screening[4,5,13-18]. Furthermore, we found a relatively high percentage of stable attenders of 56% (4073/7229) (ie, subjects attending all rounds while they were eligible). Obviously, participation depends on the willingness of participants to repeat screening. We previously observed that a positive attitude towards CRC screening, and sufficient knowledge on CRC screening are strong predictors for participation in successive rounds[19]. This suggests that increased awareness on CRC screening and sufficient information on CRC and FIT screening may have enhanced the uptake in successive rounds, as the target population was screening-naïve when first approached. In our study, previous non-attenders were re-invited. Scottish investigators reported that such practice improves uptake[16]. This is in line with our findings, where response in non-responders of previous rounds was 18.8% in round two (when inviting non-responders of the first round), and 23.2% in round three (when

inviting non-responders of the second round), thus contributing to overall participation.

The considerable decrease in DR of advanced neoplasia from 3.3% to 1.3% over the three rounds supports the notion that consecutive FIT screening has a beneficial effect by decreasing the prevalence of AN. In contrast, an Italian study on repeated FIT screening revealed a stable DR of advanced neoplasia over successive rounds (1.5-1.3%)[12]. That study however applied a higher cut-off value (100 ng Hb/ml). The initial decrease in DR in our study is likely to be explained by the enhanced sensitivity of a FIT at a low cut-off (50 ng/ml), compared to FIT screening with a higher cut-off value. In the Italian study, where average-risk subjects in the same age group as in our study were invited, the detection rate of AN was 1.5% in the first round. Even when adopting the higher cut-off of 100 ng Hb/ml as used in this study, our detection rate of AN remained higher in the first round (2.5%)[20]. This difference may be explained by the lower attendance rate during the first round than seen in our cohort (56% vs. 62%)[4,12]. Based on the above we postulate that FIT screening at a low cut-off results in a high DR and thus high sensitivity, subsequently causing a decline in DR in following rounds. In line with the decrease in DR, also a decline in PPV for advanced neoplasia was observed over repeated screening rounds. The false-positive rate (FPR), defined as subjects that had a positive FIT, but no advanced neoplasia on follow-up colonoscopy (ie, only non-advanced neoplasia, hyperplastic polyps and/or no findings at all), did not rise over the three screening rounds (FPR round 1: 5.1%; round 2: 4.1%; round 3: 4.4%, p=0.050). This indicates that the decrease in PPV was mainly due to the decrease in DR. Such a decrease in PPV in following screening rounds is what one would expect and prefer, since it is a confirmation of the effectiveness of the screening programme. The question raised based on our data, is whether the PPV has decreased too much, whether eg, the screening interval of two years that was based on the less sensitive gFOBTs used in the past, is too short. Possible ways to increase the PPV are indeed lengthening the interval, but also by using higher cut-off levels in consecutive rounds. This would of course, as a price for the higher PPV, decrease programme sensitivity and consequently the effectiveness. Evidently, there is an optimum for the PPV, where it is neither too low, nor too high. Whether a PPV for advanced neoplasia of 23% is below that optimum depends on local resources. This also needs considering the long term incidence and mortality reduction, while comparing different intensities of screening. Naturally, the models used for these analyses must be validated for whether they reproduce the low PPV in successive screening rounds as presented in this study.

Strong indicators to assess the effectiveness of a CRC screening programme are the number of screen-detected and interval CRCs over consecutive rounds. A decline in the total number of CRCs

was seen in this study, from 25 (88% screen-detected carcinomas) to 13 (54% screen-detected carcinomas). Subjects with a negative colonoscopy were referred back to the screening programme, and were considered not to require FIT screening for ten years. The reason we chose to do so and not to offer them another FIT in subsequent rounds, is because we now from previous studies that the chance of finding advanced lesions in these subjects is very low. Brenner et al. showed in a population-based case-control study that people with a previous negative colonoscopy had a strongly reduced risk of CRC compared to people who had never undergone colonoscopy[21]. Lower risks even beyond ten years after negative colonoscopy were observed for both left- and right-sided colorectal cancer, and therefore it was concluded that screening intervals for CRC screening by colonoscopy could be longer than the commonly recommended ten years in most cases[22]. A retrospective analysis found similar results, ie, that the risk of developing CRC remains decreased for more than ten years after a negative colonoscopy[23]. Findings of a more recent study also support the ten-year examination interval recommended by existing guidelines for persons at average risk who had a negative colonoscopy. Even a single negative colonoscopy was associated with a very low long-term risk of colorectal cancer[24]. However, more data are necessary to determine the optimal interval between negative colonoscopy after positive FIT and referal back to the screening programme. Another strong indicator to determine the impact of a screening programme and especially to compare it to other screening tests, is by the overall diagnostic yield of advanced neoplasia over time. For three consecutive rounds of 1-sample FIT screening, the diagnostic yield per 1,000 invitees was 43 in this study. Three rounds of FIT-based screening using a cut-off of 50 ng/ml reached a higher yield than sigmoidoscopy or colonoscopy screening, when accounting for the low uptake of these more invasive screening methods (the diagnostic yields of advanced neoplasia per 1,000 invitees of primary sigmoidoscopy and colonoscopy screening are 33 respectively 19)[7].

This study has several limitations. Recruitment took place in a population that at baseline had limited knowledge on CRC and CRC screening. Such awareness likely increased over time, particularly because of the onset of a national CRC screening programme in 2014. This may have positively affected the participation rate in the second and third screening round. Different screening intervals were applied in the second round. However, these intervals are unlikely to influence the results, since detection rates of advanced neoplasia as well as the PPV (one-year interval: 36.2%; two-year interval: 32.9%; three-year interval: 30.6%; p=0.773) were similar after a one-, two- or three-year interval, respectively[10].

In this population-based CRC screening study on three rounds of 1-sample FIT screening, an increase

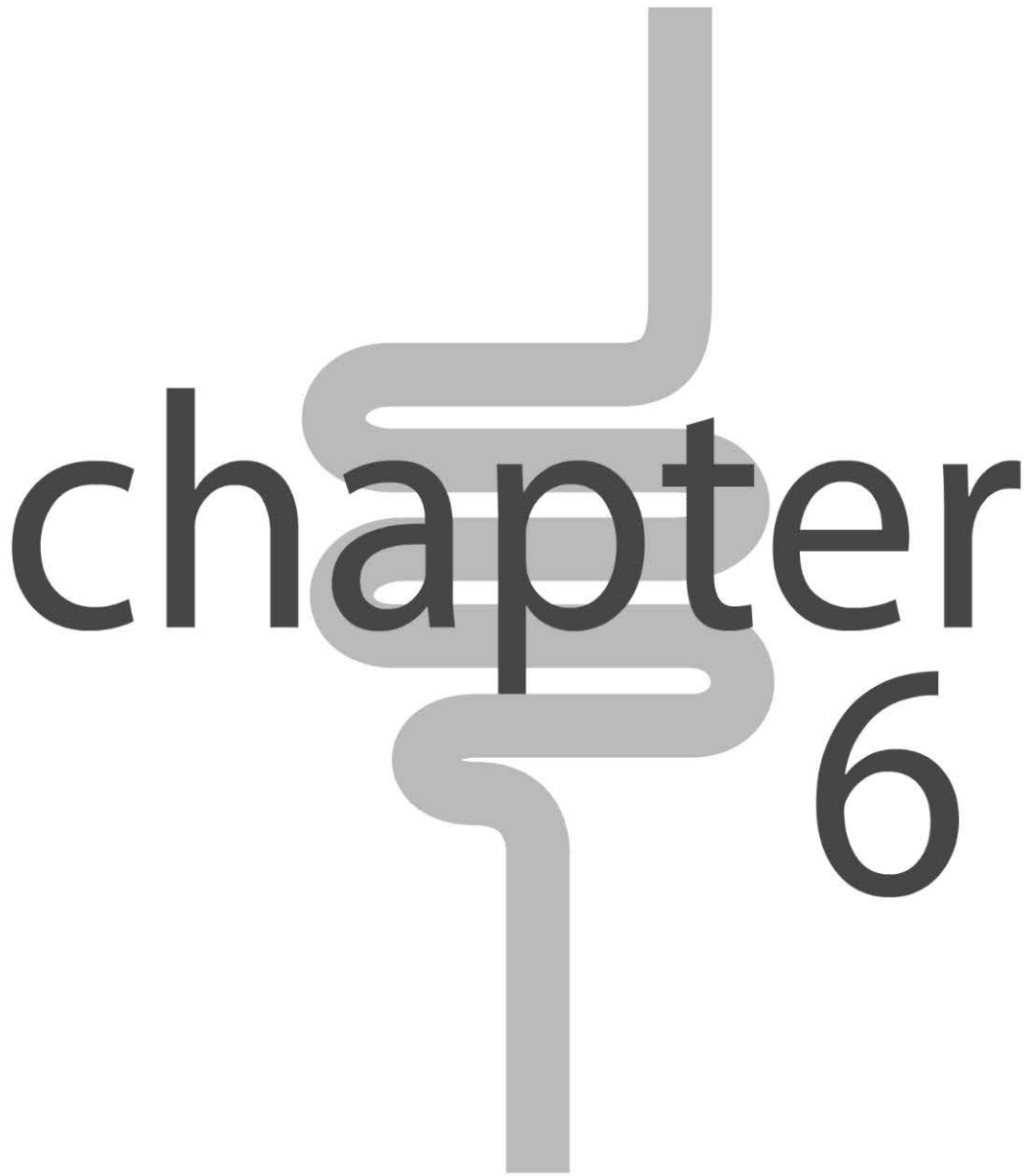
in uptake over successive screening rounds was seen. This implies that repeated FIT screening is acceptable on a population level. Furthermore, a decline in DR and PPV was seen over three consecutive rounds, suggesting that consecutive FIT screening has a beneficial effect on decreasing the prevalence of advanced neoplasia. A decrease was seen in the number of screen-detected and interval CRCs over consecutive screening rounds, providing further and even stronger evidence for the effectiveness. These results stress the importance of on-going research over multiple screening rounds. To optimize the effectiveness of screening programmes, more emphasis should be put on improving the uptake, especially that of previous non-responders.

Referenties

1. Hewitson P, Glasziou P, Irwig L, Towler B, Watson E. Screening for colorectal cancer using the faecal occult blood test, Hemoccult. Cochrane Database Syst Rev 2007; (1): CD001216.
2. Shaukat A, Mongin SJ, Geisser MS, et al. Long-term mortality after screening for colorectal cancer. N Engl J Med 2013; 369(12): 1106-14.
3. Lansdorp-Vogelaar I, van Ballegooijen M, Zauber AG, Habbema JD, Kuipers EJ. Effect of rising chemotherapy costs on the cost savings of colorectal cancer screening. J Natl Cancer Inst 2009; 101(20): 1412-22.
4. Hol L, van Leerdam ME, van Ballegooijen M, et al. Screening for colorectal cancer: randomised trial comparing guaiac-based and immunochemical faecal occult blood testing and flexible sigmoidoscopy. Gut 2010; 59(1): 62-8.
5. van Rossum LG, van Rijn AF, Laheij RJ, et al. Random comparison of guaiac and immunochemical fecal occult blood tests for colorectal cancer in a screening population. Gastroenterology 2008; 135(1): 82-90.
6. Wilschut JA, Hol L, Dekker E, et al. Cost-effectiveness analysis of a quantitative immunochemical test for colorectal cancer screening. Gastroenterology 2011; 141(5): 1648-55 e1.
7. Kuipers EJ, Rosch T, Bretthauer M. Colorectal cancer screening--optimizing current strategies and new directions. Nat Rev Clin Oncol 2013; 10(3): 130-42.
8. de Wijkerslooth TR, Stoop EM, Bossuyt PM, et al. Immunochemical fecal occult blood testing is equally sensitive for proximal and distal advanced neoplasia. Am J Gastroenterol 2012; 107(10): 1570-8.
9. Haug U, Hundt S, Brenner H. Quantitative immunochemical fecal occult blood testing for colorectal adenoma detection: evaluation in the target population of screening and comparison with qualitative tests. Am J Gastroenterol 2010; 105(3): 682-90.
10. van Roon AH, Goede SL, van Ballegooijen M, et al. Random comparison of repeated faecal immunochemical testing at different intervals for population-based colorectal cancer screening. Gut 2013; 62(3): 409-15.
11. Denters MJ, Deutekom M, Bossuyt PM, Stroobants AK, Fockens P, Dekker E. Lower risk of advanced neoplasia among patients with a previous negative result from a fecal test for colorectal cancer. Gastroenterology 2012; 142(3): 497-504.
12. Crotta S, Segnan N, Paganin S, Dagnes B, Rosset R, Senore C. High rate of advanced adenoma detection in 4 rounds of colorectal cancer screening with the fecal immunochemical test. Clin Gastroenterol Hepatol 2012; 10(6): 633-8.
13. 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.
14. Mandel JS, Bond JH, Church TR, et al. Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota Colon Cancer Control Study. N Engl J Med 1993; 328(19): 1365-71.
15. Hardcastle JD, Chamberlain JO, Robinson MH, et al. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. Lancet 1996; 348(9040): 1472-7.
16. Steele RJ, McClements PL, Libby G, et al. Results from the first three rounds of the Scottish demonstration pilot of FOBT screening for colorectal cancer. Gut 2009; 58(4): 530-5.
17. Faivre J, Dancourt V, Lejeune C, et al. Reduction in colorectal cancer mortality by fecal occult blood screening in a French controlled study. Gastroenterology 2004; 126(7): 1674-80.
18. Tazi MA, Faivre J, Dassonville F, Lamour J, Milan C, Durand G. Participation in faecal occult blood screening for colorectal cancer in a well defined French population: results of five screening rounds from 1988 to 1996. J Med Screen 1997; 4(3):

147-51.

19. van Dam L, Hol L, de Bekker-Grob EW, et al. What determines individuals' preferences for colorectal cancer screening programmes? A discrete choice experiment. Eur J Cancer 2010; 46(1): 150-9.
20. Hol L, Wilschut JA, van Ballegooijen M, et al. Screening for colorectal cancer: random comparison of guaiac and immunochemical faecal occult blood testing at different cut-off levels. Br J Cancer 2009; 100(7): 1103-10.
21. Brenner H, Chang-Claude J, Seiler CM, Sturmer T, Hoffmeister M. Does a negative screening colonoscopy ever need to be repeated? Gut 2006; 55(8): 1145-50.
22. Brenner H, Chang-Claude J, Seiler CM, Hoffmeister M. Long-term risk of colorectal cancer after negative colonoscopy. J Clin Oncol 2011; 29(28): 3761-7.
23. Singh H, Turner D, Xue L, Targownik LE, Bernstein CN. Risk of developing colorectal cancer following a negative colonoscopy examination: evidence for a 10-year interval between colonoscopies. JAMA 2006; 295(20): 2366-73.
24. Nishihara R, Wu K, Lochhead P, et al. Long-term colorectal-cancer incidence and mortality after lower endoscopy. N Engl J Med 2013; 369(12): 1095-105.


chapter
6

Adherence to colorectal cancer screening: four rounds of fecal immunochemical test (FIT)-based screening

Manon van der Vlugt, Esmée J. Grobbee, Patrick M.M. Bossuyt, Evelien Bongers, Wolfert Spijker, Ernst J. Kuipers, Iris Lansdorp-Vogelaar, Marie-Louise Essink-Bot†, Manon C.W. Spaander, Evelien Dekker

Submitted

Abstract

Background The effectiveness of fecal immunochemical test (FIT)-based screening programs is highly dependent on consistent participation over multiple rounds. We evaluated adherence to FIT-screening over four rounds and aimed to identify determinants of participation behavior.

Methods A total of 23,339 randomly selected asymptomatic persons aged 50-74 years were invited for biennial FIT-based CRC screening between 2006 and 2014. All were invited for every consecutive round, except for those who had moved out of the area, passed the upper age limit, or had tested positive in a previous screening round. A reminder letter was sent to non-responders. We calculated participation rates per round, response rates to a reminder letter, and differences in participation between subgroups defined by age, sex and socioeconomic status (SES).

Results Over the four rounds, participation rates increased significantly, from 60% (95% CI 60-61), 60% (95% CI 59-60), 62% (95%CI 61-63) to 63% (95% CI 62-64) (p for trend < 0.001) with significantly higher participation rates in women in all rounds (p<0.001). Of the 17,312 invitees eligible for at least two rounds of FIT-screening, 12,455 (72%) participated at least once, while 4,857 (28%) never participated; 8,271 (48%) attended all rounds when eligible. Consistent participation was associated with older age, female sex, and higher SES. Offering a reminder letter after the initial invite in the first round increased uptake with 12%; in subsequent screening rounds this resulted in an additional uptake of up to 10%.

Conclusion In four rounds of a pilot biennial FIT-screening program, we observed a consistently high and increasing participation rate, while sending reminders remains effective. The substantial proportion of inconsistent participants suggests the existence of incidental barriers to participation, which, if possible, should be identified and removed.

Introduction

Colorectal cancer (CRC) is a major cause of cancer-related death and its prognosis is largely dependent on stage at diagnosis[1]. Population-based CRC screening aims to detect CRC in an early stage and to detect and remove precursor lesions, thereby reducing CRC morbidity and mortality[2]. Fecal occult blood test (FOBT)-based screening using guaiac FOBT (gFOBT) has been shown to result in a reduction in CRC-related mortality in a number of randomized controlled trials with a 15% reduction in CRC-related mortality in a meta-analysis[3-6].

In the last decade, several studies have shown that the performance of the fecal immunochemical test (FIT) is superior to that of gFOBT[7-9]. Although FIT-based randomized controlled trials with long-term follow-up are lacking, a recent observational study demonstrated a 22% reduction in CRC-mortality in areas where FIT-screening programs were implemented compared to areas without screening[10]. However, FIT has a relatively low sensitivity for CRC and its precursors, and one round of FIT-screening results in a cancer miss rate of 12-25% depending on the cut-off used[11]. Screening invitations are therefore usually repeated every two years, and the effectiveness of a FIT-screening program is highly dependent on participation in multiple rounds. Ideally, eligible invitees accept the invitation to be screened in every screening round (consistent participation)[12,13].

A high rate of consistent participation increases the program-sensitivity of FIT-screening[14-17]. On the other hand, the succes of a biennial FIT-based screening program might be overestimated if the willingness to participate in multiple rounds is low. Knowing possible determinants of inconsistent participation could help in targeting the information to specific groups. Previous studies showed, for example, that especially socioeconomically deprived persons are less likely to accept CRC screening invitations[18-23].

Several studies on FOBT-screening are available, usually reporting on participation rates in a single round. We aimed to examine patterns in participation in an invitational program of biennial FIT-based screening over four screening rounds and to identify possible predictors for consistent and inconsistent screening behavior.

Methods

Study population/study design

This study was performed in our ongoing pilot program of population-based CRC screening. Details about the design of our program have been described previously[9,24-26]. In short, demographic data

of persons between 50 and 74 years living in the southwest and northwest of the Netherlands were obtained from municipal population registers. For the southwest region, random samples were taken from the target population by a computer-generated algorithm (Tenalea, Amsterdam, The Netherlands). In the northwest region random samples of selected postal code areas were taken.
The study was conducted in a dynamic cohort. Persons in the target age range that had moved into the targeted postal code area at any time during the recruitment period were included, as well as those that reached the lower age limit of 50 years.

No national screening program had been implemented at the start of this pilot-program, and thus the target population was screening-naive when first contacted. In the Netherlands, a national FIT-based CRC screening program was gradually initiated from January 2014 onwards. Invitees for our cohort were not invited for the national program. The selected persons were invited for each consecutive round, except for those who had moved out of the area, those that had passed the upper age limit, institutionalized people, invitees unable to give informed consent and those who had tested positive in a previous screening round. In our information leaflet and in our informed consent form, persons with a history of inflammatory bowel disease, proctocolectomy or CRC were asked not to participate CRC screening but report this reason for non-participation back to our screening organization via the informed consent form. Participants reporting a colonoscopy in the past 2 years during intake after a positive FIT were excluded from further participation, as well as those with an estimated life expectancy of less than 5 years. Recruitment took place between June 2006 and December 2014 (first round June 2006 to February 2007; second round August 2008 to June 2009; third round February 2011 to February 2012, fourth round March 2014 to December 2014). During the first round, invitees from the northwest region were randomly allocated to receive either a gFOBT or a FIT as screening test. Invitees who received a gFOBT in this first round were excluded from our analyses. Date of birth, sex and postal codes of all invitees were collected using the municipal population register. Socioeconomic status (SES) was based on social status scores provided by the Netherlands Institute of Social Research (www.scp.nl). The social status score of a postal code area is based on the unemployment rate, education level, average income and position on the labor market. Social status scores are available for almost all postal codes in the Netherlands. The standard deviation in the Netherlands in 2006 (start of pilot) was 0.96. Based on the social status score we assigned invitees to one of three socioeconomic status categories: high, average or low. The first available postal code of the invitee was used to categorize invitees.

FIT-screening

Every two years, all invitees received a pre-announcement letter about the screening program by mail, followed two weeks later by an invitation kit containing an invitation letter, information leaflet and a FIT with testing instructions. In the first, second and third round, all invitees received an OC-Sensor (Eiken Chemical Co, Tokyo, Japan) as FIT. In the fourth round, invitees were randomized to receive either an OC-Sensor (Eiken Chemical Co, Tokyo, Japan) or an FOB Gold (Sentinel Diagnostics SpA, Milan, Italy). As no differences in participation behavior were seen between the two tests, we included both arms in our analysis[27]. The FITs were returned to one of our two selected specialized laboratories and dates of return were registered. A test positivity threshold of ≥10 µg Hb/g feces was used. People with a positive FIT were referred for colonoscopy.

All non-responders received a reminder letter by mail after 2 to 6 weeks. Date of dispatch was registered. A positive response after the reminder letter was defined as a FIT arriving at the laboratory 3 or more days after sending out the reminder letter. This interval of 3 days was based on the mail system delivery times, which maximally take 3 days between sending and delivering. Date of dispatch of the reminder and date of return of the FIT at the laboratories were recorded for calculating return time.

Statistical analysis

The participation rate was calculated as the number of participants relative to all eligible invitees. For each screening round, we calculated participation rates per sex. For our analyses of adherence to FIT-screening, we only included invitees who were eligible in at least two rounds in order to be able to observe the three different screening patterns (see below).

Differences in screening behavior were used to assign participants to one of three groups: consistent participation (i.e. attending all rounds when eligible), inconsistent participation (i.e. attending at least once but less than the total times eligible) and non-participants (not participating in any round of FIT-screening). The percentage of consistent participants was defined as the number of invitees attending all rounds for which they were eligible relative to the total number of invitees. The percentage of inconsistent participation was defined as the number of invitees attending inconsistently relative to the total number of invitees. Similarly, the percentage of non-participants was defined as the number of invitees who never responded to any of the screening invitations.
Differences in proportions between groups were evaluated for statistical significance using the $\chi 2$-test statistic. We evaluated participation over rounds with the chi-square test statistic for trend.

Differences in medians between groups were tested using the Kruskal-Wallis test statistic. P-values <0.05 were considered to correspond to statistically significant differences. Data analysis was performed using SPSS22 for Windows (SPSS Inc., Chicago, Ill, USA).

Ethics approval

The Dutch National Health Council approved the study. All included screenees gave written informed consent to participate in the study.

Results

Population

Our dynamic cohort consisted of 23,339 invitees, of whom 323 had to be excluded because they never met the inclusion criteria; 49 invitees had moved out, 12 invitees had died and 262 invitees met one or more of the exclusion criteria (e.g having IBD, history of CRC, proctocolectomy, reporting a colonoscopy in the past 2 years during intake after a positive FIT or having an estimated life expectancy of less than 5 years) leaving 23,016 eligible invitees. Baseline characteristics of the eligible invitees are summarized in Table 1. Median age of the invitees was comparable between rounds, except for invitees in the fourth round, who had a median age of 61 (IQR 56-67). The distribution of men and women over all rounds was comparable, with 50%, 49%, 49% and 48% men, respectively (p=0.127). The majority of invitees had an average SES. The percentage of invitees with a low SES increased over the years, from 9% during the first round to 14% during the fourth screening round.

Table 1. *Baseline characteristics*

	Round 1	**Round 2**	**Round 3**	**Round 4**	**p-value**
Invitees* (n)	14,651	14,059	16,042	16,495	p<0.001
Age (median; IQR)	59 (54-65)	60 (55-65)	59 (54-65)	61 (56-67)	p<0.001
Sex (male; n, %)	7,264 (50)	6,880 (49)	7,841 (49)	7,955 (48)	p=0.127
SES (n; %)					p<0.001
Low	1,328 (9)	1,412 (10)	1,637 (10)	2,281 (14)	
Average	10,602 (72)	10,004 (71)	11,296 (70)	11,117 (67)	
High	2,721 (19)	2,643 (19)	3,094 (19)	3,088 (19)	
Missing	0	0	15 (0.1)	9 (0.1)	

*eligible invitees

Participation

Over the four rounds, participation rates increased significantly, from 60% (95% CI 60-61), 60% (95% CI 59-60), 62% (95%CI 61-63) to 63% (95% CI 62-64) respectively (Figure 1; p for trend <0.001).

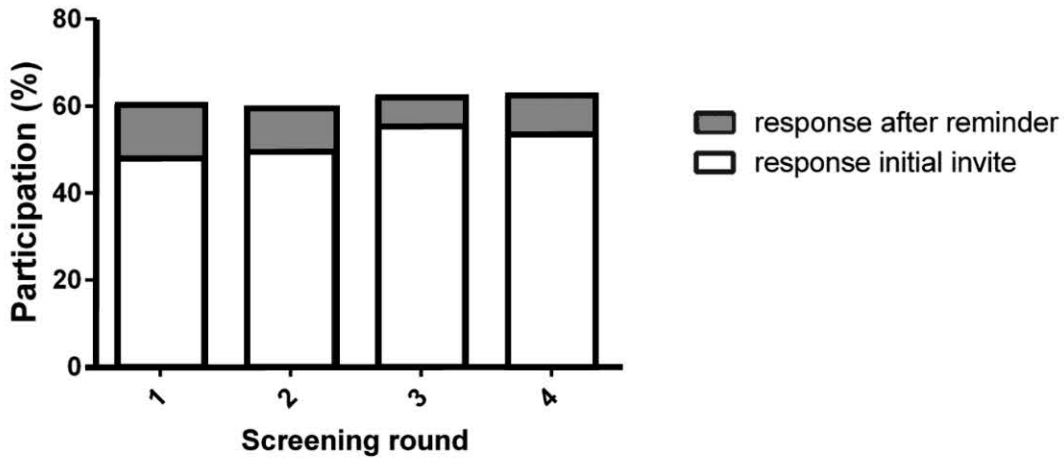


Figure 1. *Overall participation per screening round with percentage distribution of type of response to participation (initial response versus response after reminder letter).*

Differences between men and women over 4 rounds of FIT-screening are shown in Figure 2, with significantly higher participation rates for women in all 4 rounds (p<0.001).

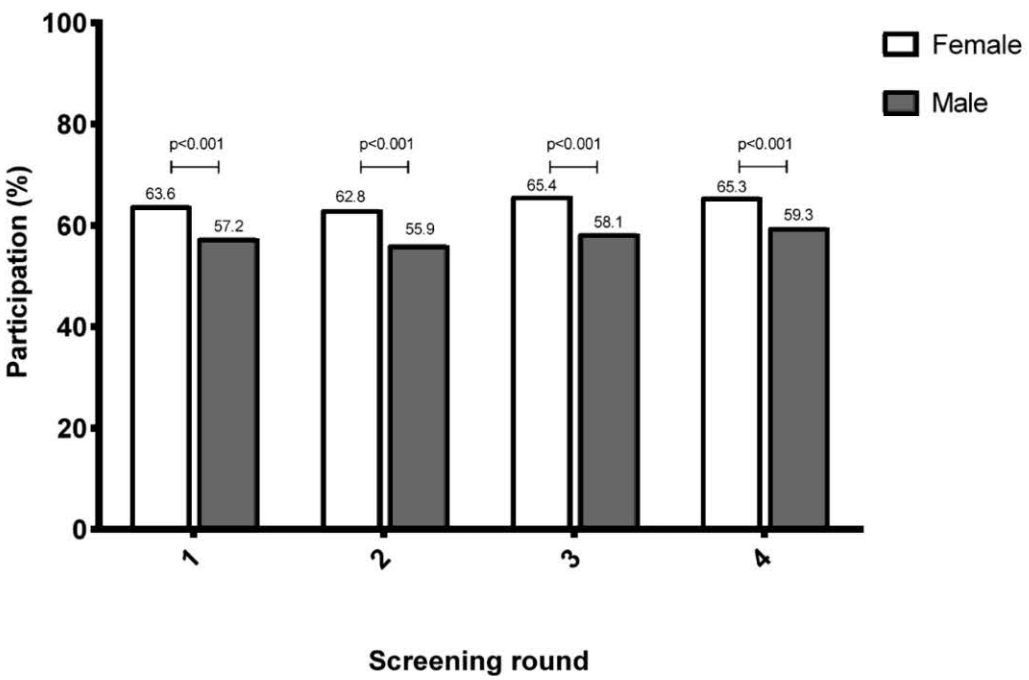


Figure 2. *Participation rates per round of FIT-based screening subdivided by sex.*

Adherence to screening and determinants of adherence

A total of 17,312 invitees were eligible for two or more rounds of FIT-screening (Table 2). In this group, 8,271 invitees (48%) were consistent participants and 4,184 (24%) were inconsistent participants.

Table 2. *Adherence to FIT based CRC screening over multiple rounds.*

	Times participated					
Times eligible	0	1	2	3	4	**total**
2	1,766 (34%)	905 (17%)	2,561 (49%)	-	-	**5,232**
3	1,011 (31%)	427 (13%)	482 (15%)	1,365 (42%)	-	**3,285**
4	2,080 (24%)	656 (8%)	686 (8%)	1,028 (12%)	4,345 (49%)	**8,795**
total	**4,857**	**1,988**	**3,729**	**2,393**	**4,345**	**17,312**

Analysis restricted to invitees who were eligible at least two screening rounds. Highlighted light grey blocks represent consistent participation

Overall, 12,455 (72%) invitees participated at least once, while 4,857 (28%) never participated in the FIT-screening program. Of the 8,795 invitees that were eligible for all four rounds, 4,345 (49%) participated in four rounds, 2,370 (27%) in one or more rounds and 2,080 (24%) participated in none. Table 3 lists the differences between consistent, inconsistent and non-participants. Consistent participants were significantly older, more often female, and more likely to have a high SES.

Table 3. *Determinants of FIT-screening participation behavior.*

	Consistent	**Inconsistent**	**Never**	**p-value**
	(n=8271, 48%)	(n=4184, 24%)	(n=4857, 28%)	
Median age * (IQR)	57 (52-63)	55 (51-61)	55 (51-62)	p<0.001
Males	45%	50%	54%	p<0.001
SES * (%)				
Low	7%	11%	14%	p<0.001
Average	71%	72%	70.%	
High	22%	17%	16%	

*age when first eligible

Reminder letter

In the first screening round, 49% (95%CI 48-49) of the invitees responded within the first 2-6 weeks after receiving the initial invitation kit, and in the first round 12% (95%CI 11-12) participated after having been sent a reminder letter (see Figure 1). The percentage of participants responding to the initial invitation increased after the first round, with participation rates of 50%, 56% and 54% for the 2nd,3rd and 4th round, respectively. An additional uptake of up to 10% was observed after sending a reminder letter (Figure 1) within every single episode. On average, FITs were returned within 15 days after sending a reminder letter (first round after 12 days (IQR 7-21); second round after 13 days

(IQR 7-32); third round after 15 days (IQR 8-27); fourth round after 14 days (IQR 7-28)).

Discussion

In four rounds of a pilot biennial FIT-screening program, we observed consistently high and increasing participation rates of 60% to 63% in each round. Sending a reminder letter after an initial non-response resulted in an increased participation rate, adding 10% to 12% in each screening round. Almost half of the invitees that were eligible for two or more screening rounds were consistent participants, while almost a quarter never participated. Consistent participants were typically older, more often female, and more likely to have a high socioeconomic status.

Strengths of our study include that our large cohort consists of an average risk population, comprising all the age ranges that are usually invited for CRC screening programs worldwide. This population was screen-naïve when first approached, without the presence of any other colorectal cancer screening initiatives in the population. More over, it covers four FIT-based screening rounds, while the majority of long term studies so far was based on gFOBT-based screening. However, some study limitations have to be acknowledged. Socioeconomic status could only be assigned by postal code, as a proxy for individual-level SES. Regrettably, no data were available on the ethnicity of all invitees, nor their marital status, both factors that could also be associated with participation[28]. Our pilot-program started in 2006, at a time that general awareness of CRC and CRC screening in the Netherlands was limited. That awareness has likely increased over time, especially after 2014, when a national Dutch CRC screening program was launched. This might have positively affected participation rates in the third and fourth screening round in our pilot program.

Similar participation rates, ranging between 56% and 63%, have been reported for a pilot study over four rounds of biennial FIT-screening in Italy[29]. Our percentage of consistent participants are in line with these data. Studies reporting on adherence to FIT-screening over a longer time interval are scarse[25,29,30]. Most reports are based on studies using gFOBT, reporting consistent adherence rates over multiple rounds, ranging from 39% to 44%[13,22,31].

As in several studies, women were more likely to participate in our FIT-screening program than men. A study from the United Kingdom also described sex differences in participation within a gFOBT-screening pilot consisting of three rounds[20]. Denis et al. reported an overall 6% higher participation rate for women in a first screening round within a gFOBT-screening program that consisted of four

rounds, with a gradually decreasing difference over time[32]. In contrast to these studies, the sex difference in our study remained comparable and significant different, though this difference was small. A possible explanation could be that women are generally more familiar with the concept of screening. In the Netherlands, women are invited for cervical cancer screening every five years, since 1996 (invitations between the age of 30-60 years), and for breast cancer screening every two years, sinds 1990 (invitations between the age of 50-75 years). So far, no other national screening programs have targeted men. Yet the fact that the difference between participation in men and women did not decrease over four rounds, in contrast to what Denis et al reported, suggests that there may be other factors involved. Possibly, men are less likely to respond the mailed invitations as compared to women and would, for instance, endorsement of the test by their general practitioner encourage them to participate.

A higher SES and older age were also significantly associated with consistent participation. These determinants for adherence to FIT-screening are comparable with those in the previously reported one-time FOBT-screening studies and gFOBT-based CRC screening studies[20-23,31,33]. Pornet et al compared occasional participants with compliant participants in a gFOBT-screening program and also reported that occasional adherence was positively assiocated with living in socioeconomically deprived areas.

Response times for participation varied over screening rounds, with prompter participation in later screening rounds. A potential explanation could be that most invitees grew familiar with the program and the FIT as a screening test over successive rounds, thereby lowering the barrier to participate and to perform the test. An alternative, additional factor could be the increased awareness of CRC and CRC screening over time.

Response rates further increased after sending reminder letters to non-participants, and this effect was seen in each of the four rounds. Previous one-time screening studies with varying intervals for sending reminder letters also showed a positive effect on uptake from sending a reminder letter[34]. Santare et al reported a very high proportion of 29% received FITs (OC-Sensor) after sending a reminder letter after 21 days, but this was studied in Latvia, which has an opportunistic screening program with very low uptake (7.6%)[35]. Tinmouth et al reported a 9.7% increase in participation after sending a reminder letter in a gFOBT-based CRC screening after 6 months. Participation rates doubled after sending a new gFOBT kit[36]. Our results indicate that sending a reminder letter to all non-responders after 6 weeks, in every screening round, consistently results in a positive

contribution to overall participation and that reminders remain effective over multiple rounds.

About one in four invitees eligible for more than one round participated once or more often, but not in all screening rounds for which they had been invited. This indicates that the decision to participate in screening is not always the outcome of a one-time assessment. It is possible that eligible citizens change their behaviour in time, and one must acknowledge that also practical issues, such as work-related responsibilities, could prevent consistent participation.

While screening uptake was high and increased over rounds, and about half of the FIT invitees were consistent participants, almost a quarter of the invitees never participated in any of the rounds of FIT-screening. It would be relevant to investigate whether these invitees made an informed decision not to participate, or whether participation was hampered by barriers, such as limited health literacy, distrust of government initiated health initiatives, cost considerations, or other issues. Health literacy is an individual's capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions. Limited health literacy has been shown to be associated with a restricted use of preventive health services, such as cancer screening[37]. A questionnaire study performed in the second round of our pilot-program of FIT-screening showed that one of the more frequently reported reasons for non-participation in FOBT-screening was lack of abdominal complaints, which suggests limitations in CRC knowledge in this group[38]. Adequate CRC knowledge was found to be a strong predictor for participation in successive rounds[39]. It is conceivable that we need to diversify our invitation and information strategy, taking into account differences between groups, to achieve equity, enabling men and women, in all age groups and socioeconomic layers, in making well-informed decisions about participation in CRC screening.

References

1. Jemal A, Bray F, Center MM, et al. Global cancer statistics. CA Cancer J Clin 2011;61:69-90 .
2. Shaukat A, Mongin SJ, Geisser MS, et al. Long- term mortality after screening for colorectal cancer. N Engl J Med 2013;369:1106-1114.
3. Hardcastle JD, Chamberlain JO, Robinson MH, et al. Randomised controlled trial of faecal- occult blood screening for colorectal cancer. Lancet 1996;348:1472–1477.
4. Kronborg O, Fenger C, Olsen J, et al. Randomised study of screening for colorectal cancer with faecal-occult-blood test. Lancet 1996;348:1467–1471.
5. Mandel JS, Bond TH, Church TR, et al. Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota Colon Cancer Control Study. N Engl J Med 1993;328:1365–1371.
6. Hewitson P, Glasziou P, Irwig L, et al. Screening for colorectal cancer using the faecal occult blood test, Hemoccult. Cochrane Database Syst Rev 2007;1:CD001216.
7. Allison JE, Sakoda LC, Levin TR, et al. Screening for colorectal neoplasms with new fecal occult blood tests: update on performance characteristics. J Natl Cancer Inst 2007;99:1462–1470.
8. Hewitson P, Glasziou P, Watson E, et al. Cochrane systematic review of colorectal cancer screening using the fecal occult blood test (hemoccult): an update. Am J Gastroenterol 2008;103:1541–1549.
9. van Rossum LG, van Rijn AF, Laheij RJ, et al. Random comparison of guaiac and immunochemical fecal occult blood tests for colorectal cancer in a screening population. Gastroenterology 2008;135(1):82–90.
10. Zorzi M, Fedeli U, Schievano E, et al. Impact on colorectal cancer mortality of screening programmes based on the faecal immunochemical test. Gut 2015;64:784-790.
11. de Wijkerslooth TR, Stoop EM, Bossuyt PM, et al. Immunochemical fecal occult blood testing is equally sensitive for proximal and distal advanced neoplasia. Am J Gastroenterol 2012;107:1570–1578.
12. Gellad ZF, Stechuchak KM, Fisher DA, et al. Longitudinal adherence to fecal occult blood testing impacts colorectal cancer screening quality. Am J Gastroenterol 2011;106:1125-1134.
13. Steele RJ, McClements PL, Libby G, et al. Patterns of uptake in a biennial faecal occult blood test screening programme for colorectal cancer. Colorectal Dis 2013;16:28-32.
14. Launoy G, Duffy SW, Prevost TC, Bouvier V. Detection of cancer, sensitivity of the test and sensitivity of the screening program. Rev Epidemiol Sante Publique 1998;46:420-426.
15. Nishihara R, Wu K, Lochhead P, et al. Long-term colorectal-cancer incidence and mortality after lower endoscopy. N Engl J Med 2013;369:1095-1105.
16. Winawer SJ, Zauber AG, Ho MN, et al. The National Polyp Study Workgroup. N Engl J Med 1993:329;1977-1981.
17. Zauber AG, Winawer SJ, O'Brien MJ, et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. N Engl J Med 2012;366;687-696.
18. Leuraud K, Jezewski-Serra D, Viguier J, Salines E. Colorectal cancer screening by guaiac faecal occult blood test in France: evaluation of the programme two years after launching. Cancer Epidemiol 2013;37:959–967.
19. Lo SH, Halloran S, Snowball J, et al. Colorectal cancer screening uptake over three biennial invitation rounds in the English bowel cancer screening programme. Gut 2015;64:282–291.
20. Moss SM, Campbell C, Melia J, et al. Performance measures in three rounds of the English bowel cancer screening pilot. Gut 2012;61:101-107.

21. Pornet C, Dejardin O, Morlais F, et al. Socioeconomic determinants for compliance to colorectal cancer screening. A multilevel analysis. J Epidemiol Community Health 2010;64:318-324.
22. Pornet C, Denis B, Perrin P, et al. Predictors of adherence to repeat fecal occult blood test in a population-based colorectal cancer screening program. Br J Cancer 2014;111:2152-2155.
23. Wee CC, Mc Carthy EP, Philips RS. Factors associated with colon cancer screening: the role of patient factors and physician counseling. Prev Med 2005;41:23-29.
24. Denters MJ, Deutekom M, Bossuyt PM, et al. Involvement of previous non-participants cannot fully compensate for lower participation in a second round of FIT-screening. Cancer Epidemiol 2013;37:330-335.
25. Kapidzic A, Grobbee EJ, Hol L, van Roon, et al. Attendance and yield over three rounds of population-based fecal immunochemical test screening. Am J Gastroenterol 2014;109:1257-1264.
26. Stegeman I, van Doorn SC, Mundt MW, et al. Participation, yield, and interval carcinomas in three rounds of biennial FIT-based colorectal cancer screening. Cancer Epidemiol 2015;39:388–393.
27. Grobbee EJ, van der Vlugt M, van Vuuren AJ, et al. Comparison of OC-Sensor and FOB-Gold in Population-Based Colorectal Cancer Screening Based on FIT. Gut 2016 Aug 9. pii: gutjnl-2016-311819.
28. El-Haddad B, Dong F, Kallail KJ, et al. Association of marital status and colorectal cancer screening participation in the USA. Colorectal Dis 2015;17:108-114.
29. Crotta S, Segnan N, Paganin S, et al. High rate of advanced adenoma detection in 4 rounds of colorectal cancer screening with the fecal immunochemical test. Clin Gastroenterol Hepatol 2012;10:633-638.
30. Jensen CD, Corley DA, Quinn VP, et al. Fecal Immunochemical Test Program Performance Over 4 Rounds of Annual Screening: A Retrospective Cohort Study. Ann Intern Med. 2016;164:456-463.
31. Lo SH, Halloran S, Snowball J, et al. Predictors of repeat participation in the NHS bowel cancer screening programme. Br J Cancer 2015;112:199–206.
32. Denis B, Gendre I, Perrin P. Participation in four rounds of a French colorectal cancer screening programme with guaiac faecal occult blood test: a population-based open cohort study. J Med Screen 2015;22:76-82.
33. Hurtado JL, Bacigalupe A, Calvo M, et al. Social inequalities in a population based colorectal cancer screening programme in the Basque Country. BMC Public Health. 2015;15:1021.
34. Baron RC, Rimer BK, Breslow RA, et al. Task Force on Community Preventive Services. Client-Directed Interventions to Increase Community Demand for Breast, Cervical, and Colorectal Cancer Screening. Am J Prev Med 2008;35:S34–S55.
35. Santare D, Kojalo I, Huttunen T, et al. Improving uptake of screening for colorectal cancer: a study on invitation strategies and different test kit use. Eur J Gastroenterol Hepatol 2015;27:536–543.
36. Tinmouth J, Patel J, Austin PC, et al. Increasing participation in colorectal cancer screening: results from a cluster randomized trial of directly mailed gFOBT kits to previous non-responders. Int J Cancer 2015;136:E697-703.
37. Kobayashi LC, Wardle J, von Wagner. Limited health literacy is a barrier to colorectal cancer screening in England: evidence from the English longitudinal study of ageing. Prev Med 2014;61:100-105.
38. Denters MJ, Bossuyt PM, Deutekom M, et al. Most participate in faecal immunochemical test-based colorectal cancer screening out of curiosity about their chances of developing cancer. Eur J Cancer Prev 2015;24:176–179.
39. van Dam L, Korfage IJ, Kuipers EJ, et al. What influences the decision to participate in colorectalcancer screening with faecal occult blood testing and sigmoidoscopy? Eur J Cancer 2013;49:2321-2330.


chapter
7

First steps towards combining fecal immunochemical testing with the gut microbiome

Esmée J. Grobbee, Gwenny M. Fuhler, Blerdi Blakaj, Sergey R. Konstantinov, Marco J. Bruno, Maikel P. Peppelenbosch, Ernst J. Kuipers, Manon C.W. Spaander

Submitted

Abstract

Objectives Many countries use fecal immunochemical testing (FIT) to screen for colorectal cancer (CRC). However, FIT has a lower sensitivity for precancerous adenomas and serrated lesions. There is increasing evidence that fecal microbiota play a crucial role in CRC carcinogenesis. Therefore, we assessed the possibility of measuring the fecal microbiome in FIT as additional biomarker in CRC screening.

Methods Positive FIT samples (n=200) of an average-risk screening cohort were analyzed for universal 16S, and bacteria previously associated with CRC, i.c. Escherichia coli, Fusobacterium nucleatum, Bacteroides and Fecalibacterium prausnitzii by qPCR. The results were compared to colonoscopy findings.

Results Of the FIT-positive samples, 20 had to be discarded for various reasons resulting in 180 samples. Fecal microbiome was stably measured with no significant decrease in fecal microbiota up to 6 days for E. coli (p=0.53), F. nucleatum (p=0.30), Bacteroides (p=0.05) and F. prausnitzii (p=0.62). Total bacterial load (i.e. 16S) was significant higher in patients with CRC and high-grade dysplasia (p=0.006). For individual bacteria, relative to 16S, no association was found with colonic lesions. No significant differences in fecal microbiome in FIT were found between screenees with and without advanced neoplasia.

Conclusions To the best of our knowledge we are the first to describe the use of measuring microbial content in FIT samples for population-based CRC screening. Our results show that the fecal microbiome can be measured in FIT samples and remains stable for 6 days. Total bacterial load was higher in CRC and high-grade dysplasia. These results pave the way for further research to determine the role of microbiota in FIT, to increase FIT sensitivity.

Introduction

Colorectal cancer (CRC) is a major cause of cancer-related morbidity and mortality 1. Its burden is likely to increase by 60%, to over a million cancer deaths by 2030 2. The etiology of CRC is complex and not yet completely understood. It is believed that, in addition to hereditary factors, environmental risk factors such as a Western diet, smoking, lack of physical activity, and obesity play an important role in the development of CRC 3. Furthermore, there is increasing attention for the gut microbiome and its role in colorectal carcinogenesis[4-6]. It is estimated that at least 20%, and perhaps more, of the cancer burden worldwide can be attributed to microbial agents[7].

An association between CRC and specific fecal bacteria was already reported a long time ago[8]. In a small Dutch study of 12 patients with Streptococcus bovis bacteremia, we diagnosed CRC in eight and gastric cancer in one patient[9]. In contrast to gastric cancer, in which a single pathogen plays a predominant carcinogenic role, CRC appears to have a more complex etiology with potential etiological contribution of multiple bacterial species playing different roles[6,10-12]. Most gut bacteria cannot easily be cultivated, yet sequencing of bacterial DNA using quantitative polymerase chain reaction (qPCR) nowadays allows identification of the composition of the fecal microbiota[6]. Evidence to date suggests that inflammatory processes triggered by enterotoxigenic bacteria (i.e. alpha-bugs) can contribute to CRC development by facilitating DNA damage in intestinal epithelial cells (IEC)[8]. Bacteria such as Bacteroides and Enterobacteriaceae have been shown to induce pro-inflammatory factors and reactive oxygen production in IEC[11,13]. The ensuing accumulation of genetic lesions can contribute to oncogenesis along the adenoma-carcinoma sequence. Several studies have shown that the bacterial composition of malignant lesions differs from that of surrounding normal tissue[14-16]. This resulted in the postulation of the so-called 'driver-passenger model'[11]. This theory dictates that pro-oncogenic driver bacteria, (the alpha-and helper bugs) help trigger the development of colorectal neoplasia, while disease progression causes an altered microenvironment that favors other (commensal) bacteria. Such 'passenger' bacteria, which may or may not promote tumor growth, can eventually out-compete the 'driver' bacteria[11,17]. This theory suggests that fecal microbiota is dynamic during the course of disease progression, and that different (pre)malignant lesions may be associated with different modulations of the microbiome. While most previous research has focused on the role of the gut microbiome in the pathogenesis of CRC, it is of significant interest to see whether altered bacterial presence may be valuable in improving screening strategies for CRC[15,18].

In the past decennia an increasing number of countries have embarked on CRC screening. Many

of those use fecal immunochemical tests (FIT) as their screening method[19]. FIT tests rely on the measurement of trace amounts of blood from neoplastic lesions. However, not all lesions bleed (in particular serrated adenomas are less prone to blood loss), and conversely, occult blood can be detected in fecal samples of healthy individuals[20]. In spite of high participation rates and a relatively high sensitivity for CRC of 75-85% depending on the cut-off used, the sensitivity of FIT for detection of advanced adenomas is much lower and generally ranges below 50%[21-23]. For this reason there is an urgent need for additional markers to increase FIT sensitivity without losing its specificity, as the latter is of crucial importance in a screening setting. Investigation of the fecal bacterial composition could present one such possible additional marker. Hence, it would be of great interest to detect bacterial species in the rest materials of FIT-screenees, which would preclude additional material collection from screenees. To the best of our knowledge no study has yet described the detection of microbiota in FIT. Therefore, the aim of our study was to evaluate the possibility of measuring fecal microbiota in FIT, and to assess the potential use of bacterial species detection as an additional diagnostic biomarker in CRC screening. Based on their previously described association with CRC, we selected bacteria from 4 different genera: suspected driver bacteria of the Enterobacteriaceae (Escherichia coli) and Bacteroides species, the most often associated CRC bacterium, Fusobacterium nucleatum, and the anti-inflammatory Clostridiaceae Fecalibacterium prausnitzii (F. prausnitzii) which was described to be less prevalent in CRC patients[18,24].

Methods

Patients, FIT-screening, and data collection

Details about the design of this ongoing population-based CRC screening program have been described previously[25,26]. In short, demographic data of all individuals between 50 and 74 years living in the southwest of the Netherlands were obtained from municipal population registers. Random samples were taken from the target population by a computer-generated algorithm (Tenalea, Amsterdam, The Netherlands). As there was no CRC screening program at the time of the trial in the Netherlands, the target population was screening-naive when first approached. At present, four rounds of FIT-screening have taken place. Individuals with a history of inflammatory bowel disease or CRC, as well as those who had undergone a colonoscopy, sigmoidoscopy, or barium contrast enema in the past 3 years, those with an estimated life expectancy of <5 years, and subjects who were unable to give informed consent were excluded from the study. For this study only FIT samples of the end of the third and beginning of the fourth screening round were used, aiming for a total of 200 FITs to be included. Recruitment of this third and fourth screening round took

place between February 2013 and August 2014. In the third screening round all invitees received the OC-sensor (Eiken, Japan). In the fourth screening round invitees were randomized between the OC-sensor and FOB-Gold (Sentinel, Italy). Participants were instructed to send the FIT within one day after collection and to keep the FIT in the refrigerator until sending it to the laboratory. A cut-off of ≥10 µg Hb/g feces was used to refer the screenee for a colonoscopy within 4 weeks. All colonoscopies were performed by gastroenterologists with an experience based on at least 1,000 colonoscopies. All colonoscopy findings were prospectively registered using pre-defined clinical registration forms. All lesions were evaluated by trained gastrointestinal pathologists according to the Vienna criteria[27]. Advanced adenomas (AA) were defined as an adenoma with a diameter ≥10 mm, and/or with a ≥25% villous component, and/or high-grade dysplasia. Advanced neoplasia (AN) included AA and CRC, with the most advanced lesion used for analysis. Serrated polyps were defined as serrated adenomas (with our without dysplasia) and hyperplastic polyps. For this study only FIT-positive screenees were included.

Bacterial quantitative analysis

After occult blood measurement, FIT samples were stored at -20°C until analysis. DNA was isolated from FIT liquid by Wizard Genomic DNA Purification kit (Promega, Leiden, the Netherlands) with modifications. Bead-beating was performed 3 times for 30 seconds to lyse bacteria. Protein was precipitated from the supernatant by Protein Precipitation Buffer, followed by isopropanol precipitation of DNA. DNA was washed with 70% ethanol and resuspended in Tris-EDTA (TE) buffer. DNA concentration was measured by Nanodrop (ThermoFisher Scientific), and adjusted to 10ng/µL. Bacterial DNA (E. coli, F. nucleatum, Bacteroides, F. prausnitzii and universal bacterial 16S) was detected by PCR or qPCR. Specificity of primers (Table 1) was determined by primer blast search against all bacteria (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) 28-37. In addition, DNA was amplified by PCR using GoTaq polymerase (Promega) and 35 cycles (95° for 15", 56° for 30", 72° for 30") on a 2720 Thermal Cycler (Applied Biosystems, ThermoFisher) and PCR products were verified by 2% agarose gel-electroforesis (for examples see Figure 1). Analysis of bacterial abundance was performed by SybrGreen based quantitative PCR (qPCR). qPCR reactions were performed in a 20µL volume, containing 10µL SYBR® Select Master Mix for CFX (ThermoFisher Scientific), 2µL Forward+ Reverse primer (end concentration 1µM), 7µL H2O and 1µL template. Cycle conditions were 95° for 10 min (initial denaturation), followed by 40 cycles of (95° for 15", 56° for 30", 72° for 30") and melt curve analysis on StepOnePlus real time PCR System (Applied Biosystems, Thermofisher).

For each bacterium, DNA for standard curves was prepared by PCR using DNA isolated from FIT tests as template. PCR products were purified using Invisorb® Fragment CleanUp kit (Stratec molecular,

Berlin, Germany), DNA concentration was measured, and DNA dilutions ranging from 0.0001-10 ng/µl were prepared. The amount of DNA in test samples was inferred from their Ct value through calculation of standard curves run on each separate plate, and corrected for total amount of DNA present in the FIT liquids. Copy number per gram PCR product was inferred from the weight of one PCR product as calculated by the accumulated weight of the basepairs in the product, and results were presented as absolute bacterial content (copy number/g FIT fluid) or as ratio of the total amount of 16S copies/gr FIT fluid. Samples where bacterial DNA was undetectable were given a value <than the lowest detectable value for that PCR product, in order to prevent loss of these samples from group comparisons.

Table 1. *Primer sequences used in this study*

Bacterium		Sequence (5'-3')	Product size	References
Universal 16S	Bact-1369-F	cggtgaatacgttcccgg	145	28,38
	Bact-1492-R	tacggctaccttgttacgactt		
F. nucleatum	Forward	cttaggaatgagacagagatg	140	29,30
	Reverse	tgatggtaacatacgaaagg		
E. coli	Forward	catgccgcgtgtatgaagaa	96	31,32
	Reverse	cgggtaacgtcaatgagcaaa		
Bacteroides	Forward	cggacgtaagggccgtgc	140	33,34
	Reverse	ggtgtcggcttaagtgccat		
F. prausnitzii	Fprau223F	gatggcctcgcgtccgattag	198	35,36
	Fprau420R	ccgaagaccttcttcctcc		

Microbiota stability in FIT

For analysis of the stability of the microbial content of FIT over time, seven FIT tests were taken from one stool sample of a healthy volunteer, and tests were stored at -20°C immediately, or after 24, 48, 72, 96, 120 and 144 h in order to mimic FIT transit time. The experiment was performed three times, with the same fecal donor. DNA was isolated from all samples simultaneously upon thawing, and presence of E. coli, F. nucleatum, Bacteroides, F. prausnitzii and universal bacterial 16S were detected by PCR and qPCR as described above. F. nucleatum and E. coli were below detection level in this donor. Therefore, stability of bacterial DNA was also tested by spiking tests with a known concentration (0.012 ng/µl) of E. coli PCR product, which has the added benefit of circumventing measurement of bacterial growth/death. Spiked FIT without stool was used as control.

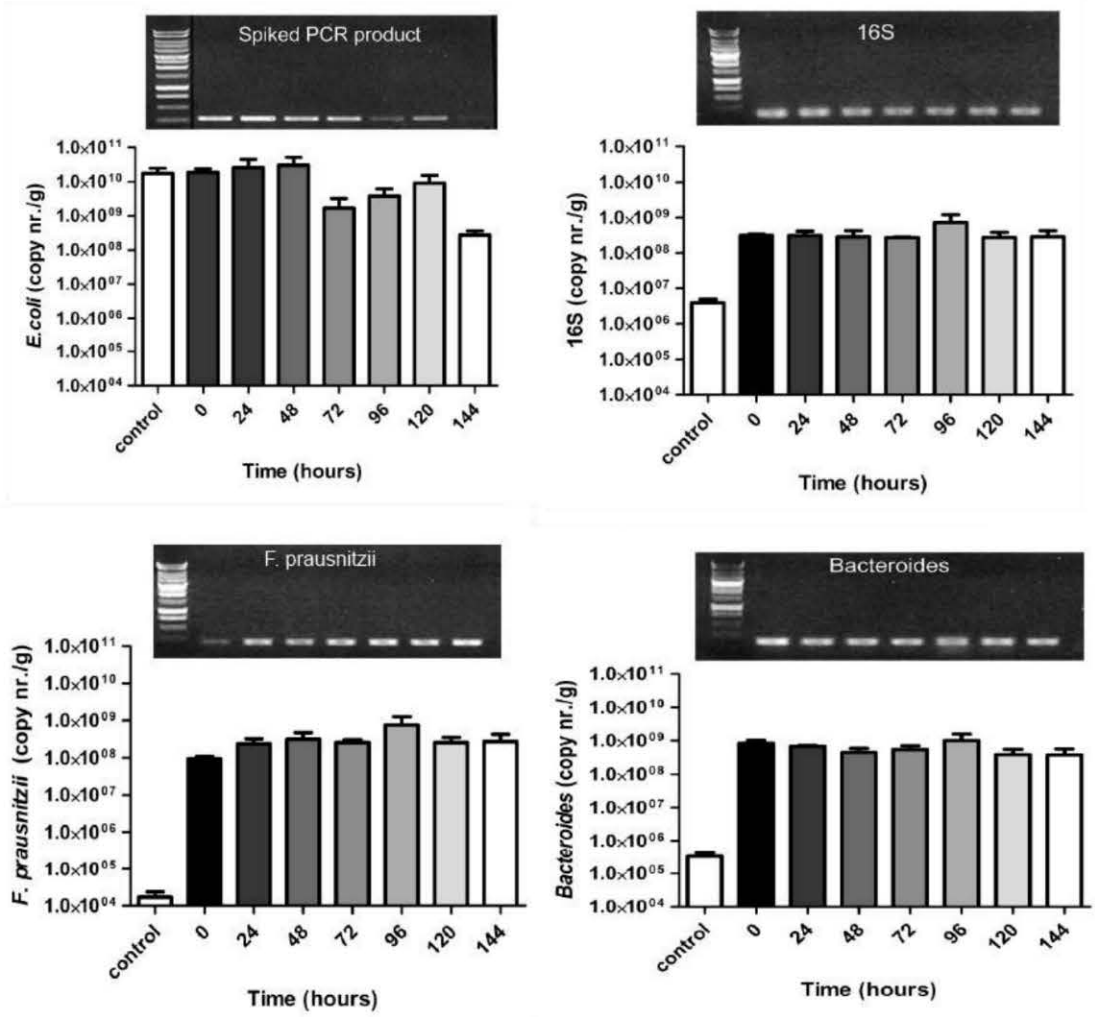


Figure 1. *Stability of bacterial composition in FIT in spiked E. coli PCR (A), 16S (B), F. prausnitzii (C), Bacteroides (D). Bacterial composition remained stable up to at least 144 hours.*

Statistical analysis

Descriptive data were reported as proportions or means with the standard deviation. For non-normally distributed data the median and interquartile range (IQR) were given. Chi-Square tests were used to analyze categorical data; continuous data were analyzed using Student's t-tests or one-way ANOVA. Linear regression analysis was used to assess bacterial load and transit time. Correction for multiple testing was done according to Bonferroni resulting in a two-sided p-value of <0.01 that was considered to be statistically significant. Statistical analysis was performed using IBM SPSS version 21.0.

Ethical approval

The study was approved by the Medical Ethical Committee of the Erasmus MC, Rotterdam (reference number: MEC-2014-212).

Results

FIT screenees

A total of 200 samples from FIT positive screenees were collected. Of these, 20 samples had to be discarded for various reasons (e.g. because multiple samples from the same screenee were included, because a sample was misclassified, or because pathology outcome was missing), resulting in 180 samples available for analysis of microbial content. Of those, 56% were male with a median age of 64 years (IQR 58-69 years). Median fecal Hb concentration was 21 µg Hb/g feces (IQR 13-55 µg Hb/g feces). All screenees included in this study underwent colonoscopy and in 31% (n=55) patients advanced neoplasia was detected, of whom 5 were diagnosed with CRC. All colonoscopy findings are described in Table 2.

Table 2. *Most advanced lesion at colonoscopy of FIT positive screenees*

Finding at colonoscopy	n (%)
Normal	41 (22.4)
Serrated polyps	25 (13.7)
Tubular adenoma < 10 mm	59 (32.2)
Tubular adenoma >=10 mm	33 (18.0)
(Tubulo)villous adenoma	14 (7.7)
High-grade dysplasia	3 (1.6)
Colorectal carcinoma	5 (3.3)
Total	180 (100*)

*numbers do not add up exactly to 100% due to rounding

Stability microbiota in FIT over time

Transit time of the FIT from screenee to the laboratory could potentially affect the microbial composition detected. Although growth of anaerobic bacteria is not expected, and FIT buffer contains bacteriostatic sodiumazide, degradation of bacterial DNA might occur. Therefore, we first analyzed the stability of the bacterial composition in FIT. An artificially spiked E. coli PCR product could be reliably measured in stool samples that were left at ambient temperature for up to 48h (Figure 1A).

Endogenous universal bacterial 16S, F. prausnitzii and Bacteroides DNA was consistently detected by (q)PCR for even longer periods, with no loss in detection levels for up to 144h (Figure 1B-D).

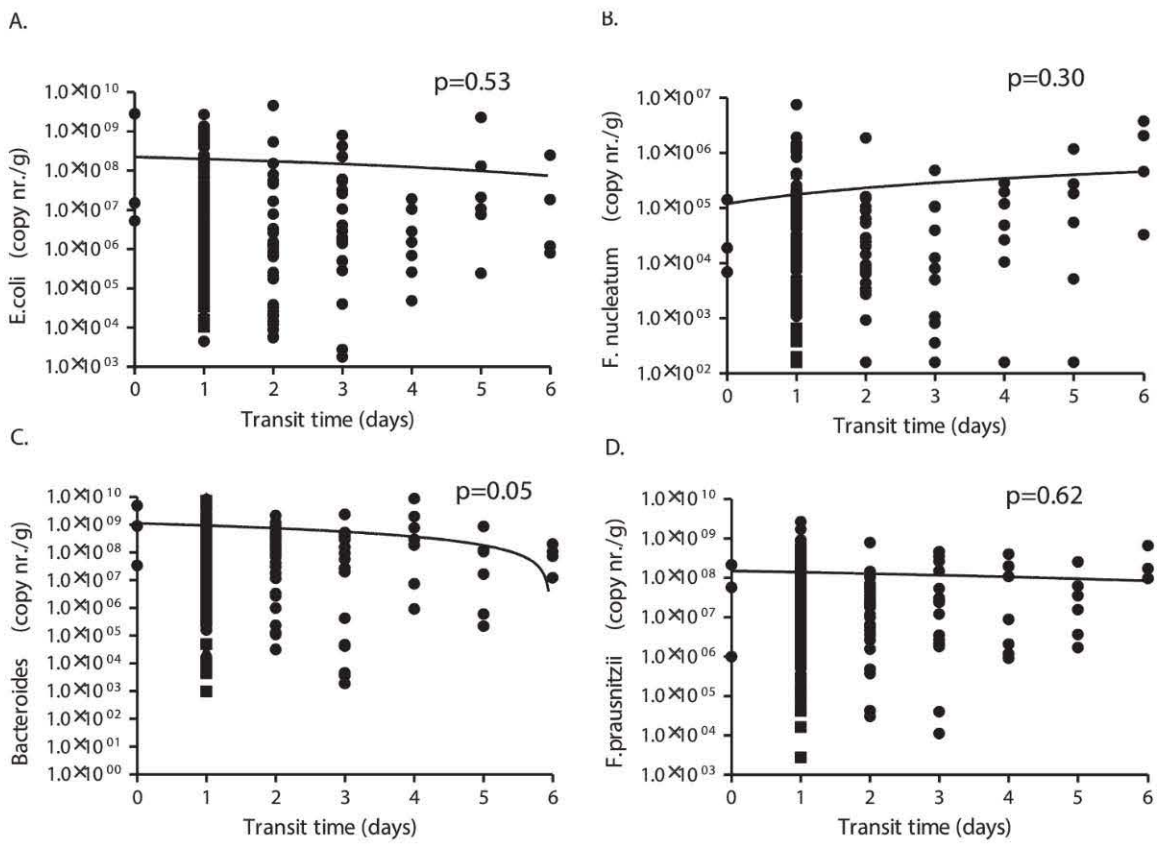


Figure 2. Transit time (interval between fecal sampling and arrival of the FIT specimen at the laboratory) and absolute copy number/gram FIT test per 16S for E. coli (A), F. nucleatum (B), Bacteroides (C), F. prausnitzii (D).

The average time between fecal sampling by the screenee and analysis at the laboratory (i.e. transit time) was 1 day (IQR 1-2 days), with 91% of FITs arriving at the laboratory within 2 days after sampling. For all screenees the correlation between absolute copy number of the four bacteria and transit time were evaluated (Figure 2). No significant decrease in fecal microbiota was seen up to 6 days for E. coli (p=0.53), F. nucleatum (p=0.30), Bacteroides (p=0.05) and F. prausnitzii (p=0.62, Figure 2).

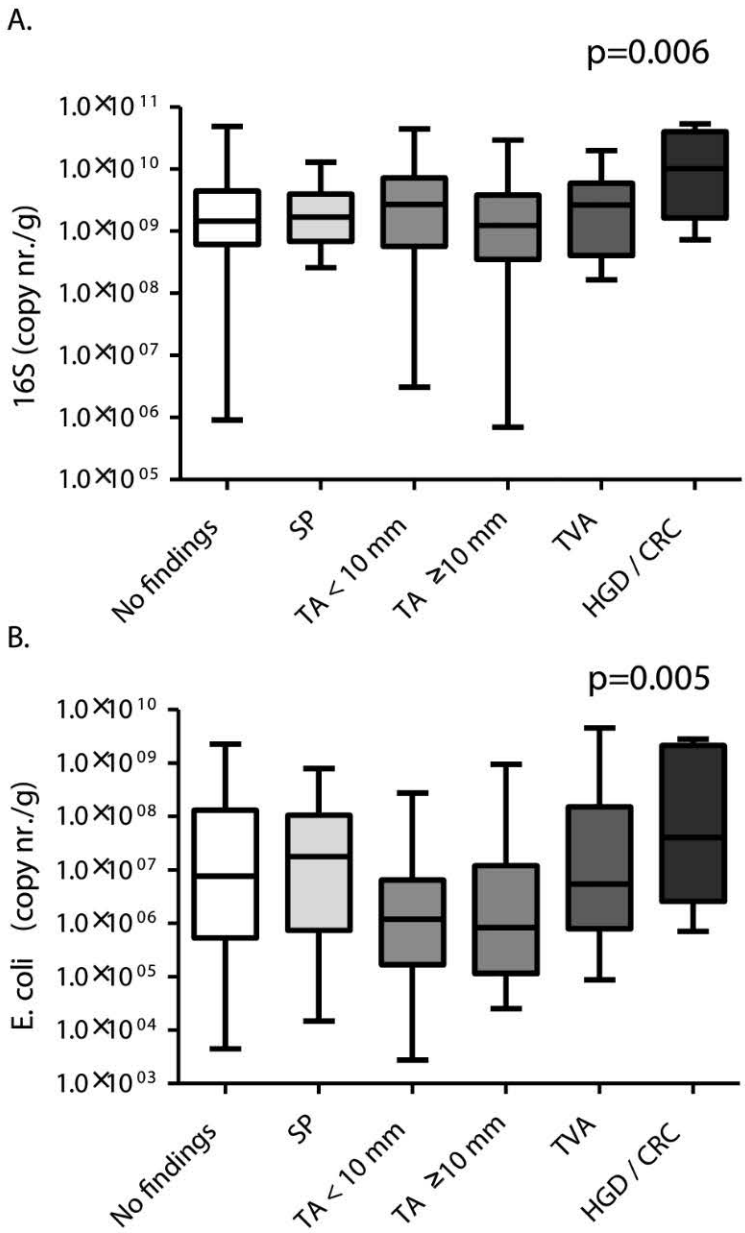


Figure 3. *Copy number per gram FIT liquid for 16S (A) and E. coli (B) according to colonoscopy outcomes.*

Microbiome in FIT and findings at colonoscopy

For all samples, copy number per gram (copy nr./g) FIT liquid was calculated for the total number of bacteria (i.e. 16S) and the four predefined bacteria. A significant difference was seen for 16S , with increasing abundance of total bacterial content in screenees with high-grade dysplasia and CRC (p=0.006; Figure 3A). For E. coli, the copy nr./g was lower in patients with tubular and villous adenomas compared to patients with a normal colonoscopy, serrated polyps, high-grade dysplasia and CRC (p=0.005; Figure 3B). For F. nucleatum, F. prausnitzii, and Bacteroides, no association was

observed between the presence of the bacteria and any particular lesion (Supplementary Figure 1). No significant association between amounts of bacteria and presence of advanced neoplasia was observed (Supplementary Figure 2).

To correct for potential differences in amount of fecal matter in the FIT, the bacteria were also calculated relative to the total bacterial presence as determined by universal 16S (copy nr./g of 16S). No significant differences were found when evaluating FIT microbiota according to all colonoscopy findings, including CRC, for E. coli (p=0.97), F. nucleatum (p=0.98), Bacteroides (p=0.15) and F. prausnitzii (p=0.91; Figure 4).

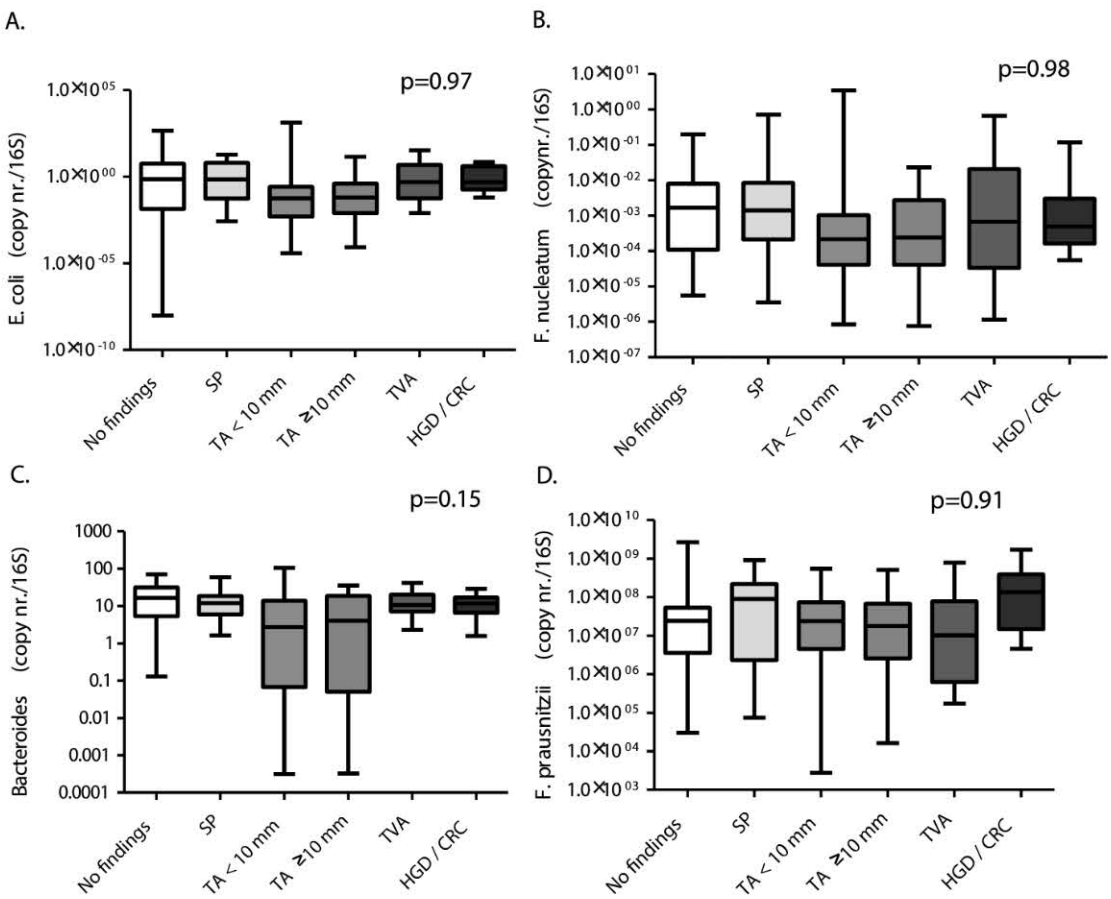


Figure 4. *Absolute copy number per 16S and most advanced colonoscopy finding* for E. coli (A), F. nucleatum (B), Bacteroides (C) and F. prausnitzii (D).*

In addition, no significant differences in microbiota were found between screenees with and without advanced neoplasia (E. coli p=0.30; F. nucleatum p=0.55; Bacteroides p=0.12; F. prausnitzii p=0.93; Figure 5). When evaluating FIT microbiota according to location of the most advanced lesion (i.e.

distal vs. proximal), again no significant differences were seen for all four bacteria (Supplementary Figure 3).

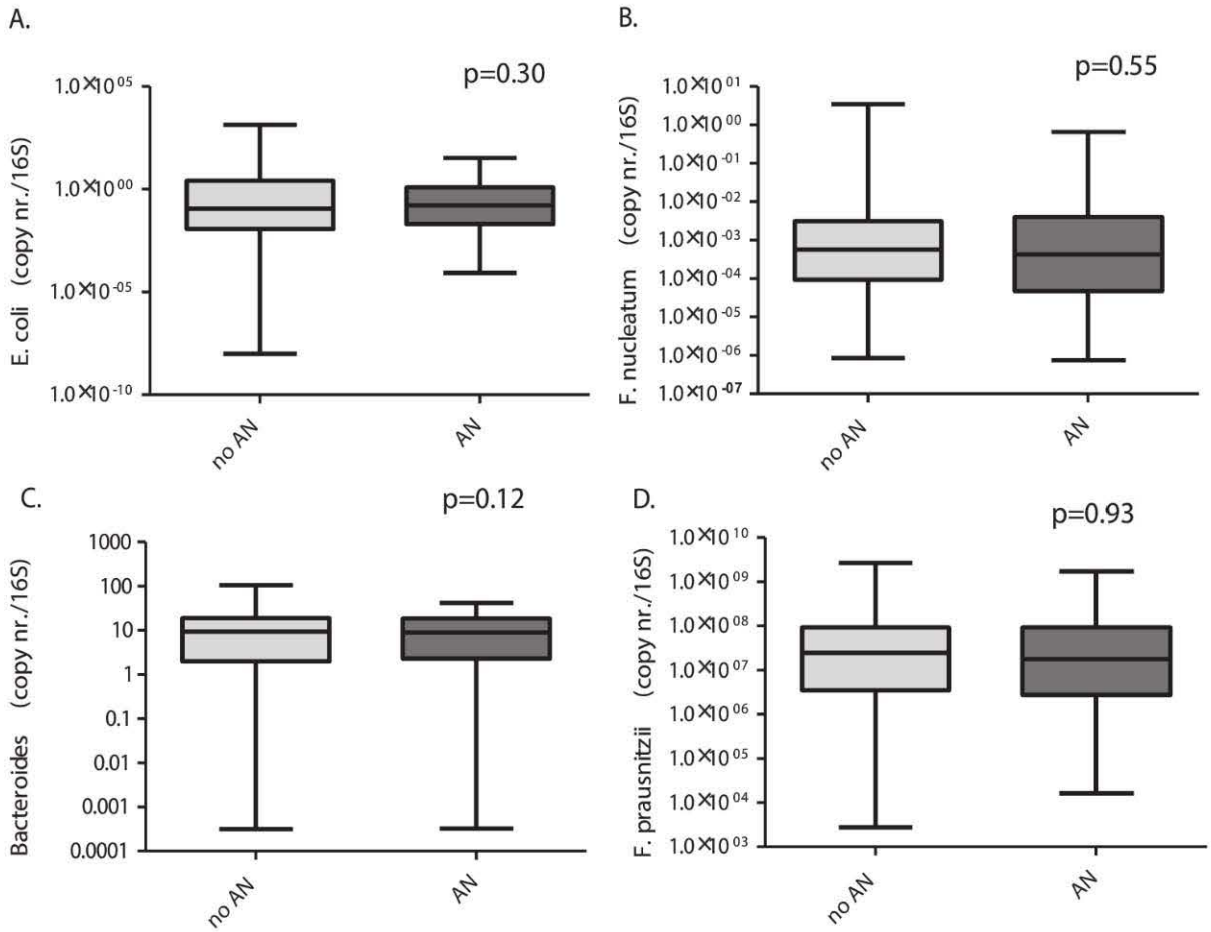


Figure 5. *Microbiome in absolute copy number per 16S for screenees with no advanced neoplasia (no AN) and advanced neoplasia (AN) for E. coli (A), F. nucleatum (B), Bacteroides (C), F. prausnitzii (D).*

Discussion

To the best of our knowledge we are the first to describe assessment of the microbiome in FIT samples. Our results show that fecal microbial DNA can be isolated from FIT samples and that it remains stable up to 6 days. Screenees with high-grade dysplasia and CRC had a higher load of total universal 16S. With respect to specific microbial composition, no relation was found between numbers of specific bacteria and colonoscopy findings relative to total 16S, except that numbers of E. coli were reduced in patients with tubular and villous adenoma. With regard to location of the lesion no differences were found between a lesion in the distal or proximal colon and number of fecal bacteria found. For this study four bacterial species were chosen, together representing a

wide spectrum of genera thought to be of importance in CRC carcinogenesis. E. coli and Bacteroides species are considered to promote inflammation driving the colorectal epithelium to a carcinogenic state[6,18,39]. This state of inflammation and dysbiosis gives room for opportunistic bacteria, such as F. nucleatum, to further induce carcinogenesis, whereas anti-inflammatory bacteria such as F. prausnitzii may be 'crowded out' [24].Consequently various bacteria take part in the process of carcinogenesis, with many of these bacteria being variably present during carcinogenesis[11]. This could explain why lower concentrations of E. coli were found in tubular and villous adenomas, although screenees with normal colonoscopy and those with CRC had similar concentrations. While E. coli and Bacteroides have been identified in a number of studies as carcinogenic factors in CRC, our findings did not support a role for these bacteria as additional biomarker in FIT samples to identify FIT-positive screenees at risk of carrying advanced adenomas.

A previous study has suggested a role for F. nucleatum in the detection of sessile serrated lesions, with the mucus cap on these lesions suggested as a cause for the high levels of F. nucleatum[14]. Additional detection of sessile serrated lesions would be especially valuable in FIT screening, as FIT is known to have a poor sensitivity for these lesions[40]. However, our results did not show any association between F. nucleatum and hyperplastic polyps or serrated lesions compared to other neoplasia or a normal colon (p= 0.82; data not shown). This could be because F. nucleatum is not sensitive enough by itself as a biomarker in a screening setting due tooverabundance in healthy subjects[13]. Further studies have to assess the prevalence of F. nucleatum in FIT-negative screenees who also undergo colonoscopy.

Most studies regarding the role of the microbiome in colorectal carcinogenesis have looked specifically at the microbiome at and around the tumor-site and it is conceivable that a fecal sample obtained by FIT is not representative of onsite mucosal dysbiosis[41]. However, microbiota on mucosa retrieved during colonoscopy or surgery could be influenced by the bowel preparation that all patients undergo prior to the intervention. Furthermore, most of these studies had a case-control design and were thus prone to overestimate diagnostic performance[42]. To date, there are two other studies that looked at the fecal microbiome in FIT screenees, showing a difference in overall fecal microbiome between normal patients and patients with colorectal adenomas[15,43]. In both studies the microbiome was analyzed in full stool samples and not in the FIT samples themselves, making comparison with our data complex. However, a full stool sample may ask for a considerable effort from the screenee, making the design undesirable in a screening setting as it might hamper participation rates. Furthermore, the microbiota-based model described in one of these studies was based on sequencing a selection of stool samples[43]. Such a data-driven strategy has several

limitations and can lead to overly optimistic estimates of diagnostic test accuracy[44].

The strengths of our study include the fact that all FIT samples were retrieved from a population-based CRC screening cohort, consisting of average risk screenees, resulting in a high external validity. Also, as gut microbiota were measured in FIT samples, no additional stool samples were required from the participants. It is the first study showing fecal microbiota related to colonoscopy outcomes ranging from no findings to CRC. This allows comparison of microbiota between previously untreated patients across the adenoma-to-carcinoma range. Furthermore, we included only FIT samples that tested positive for occult blood, precluding the possibility of a bias introduced by potential microbe-blood interactions[45]. However, in order to appreciate our results, some limitations also need to be addressed. Firstly, we did not compare our findings to histological material to confirm the presence of bacteria on the mucosal surface. Furthermore, as only FIT-positive subjects underwent colonoscopy, it was not possible to evaluate prime indicators of diagnostic performance, including sensitivity, specificity and the area under the receiver-operating curve. However, we considered analysis of only FITs justified, as this is representative of population-based FIT screening where screenees do not undergo colonoscopy as primary screening method. Moreover, this is in line with our aim to study the presence of microbiota in FIT samples since this would enable implementation of microbiota assessment in FIT samples as an important additional tool. Secondly, we used the most advanced lesion detected during colonoscopy, while screenees sometimes have more than one lesion. The presence of multiple lesions could theoretically lead to our findings being an underestimation of the relation between fecal microbiome and colonic neoplasia. Yet, for screening purposes subjects at highest risk (i.e. with advanced neoplasia) are of most interest and therefore we believe that our findings adequately represent screening outcomes. Also, distribution of lesions is in accordance with previous literature on diagnostic yield of repeated rounds of FIT screening[25,46].

In conclusion, our results illustrate that the gut microbiome can be measured in FIT-samples in CRC screening, with a higher total bacterial load for CRC and high-grade dysplasia. The need to increase FIT sensitivity, especially for advanced adenomas, remains of evident importance and further studies should be conducted to determine the role of microbiota in FIT.

References

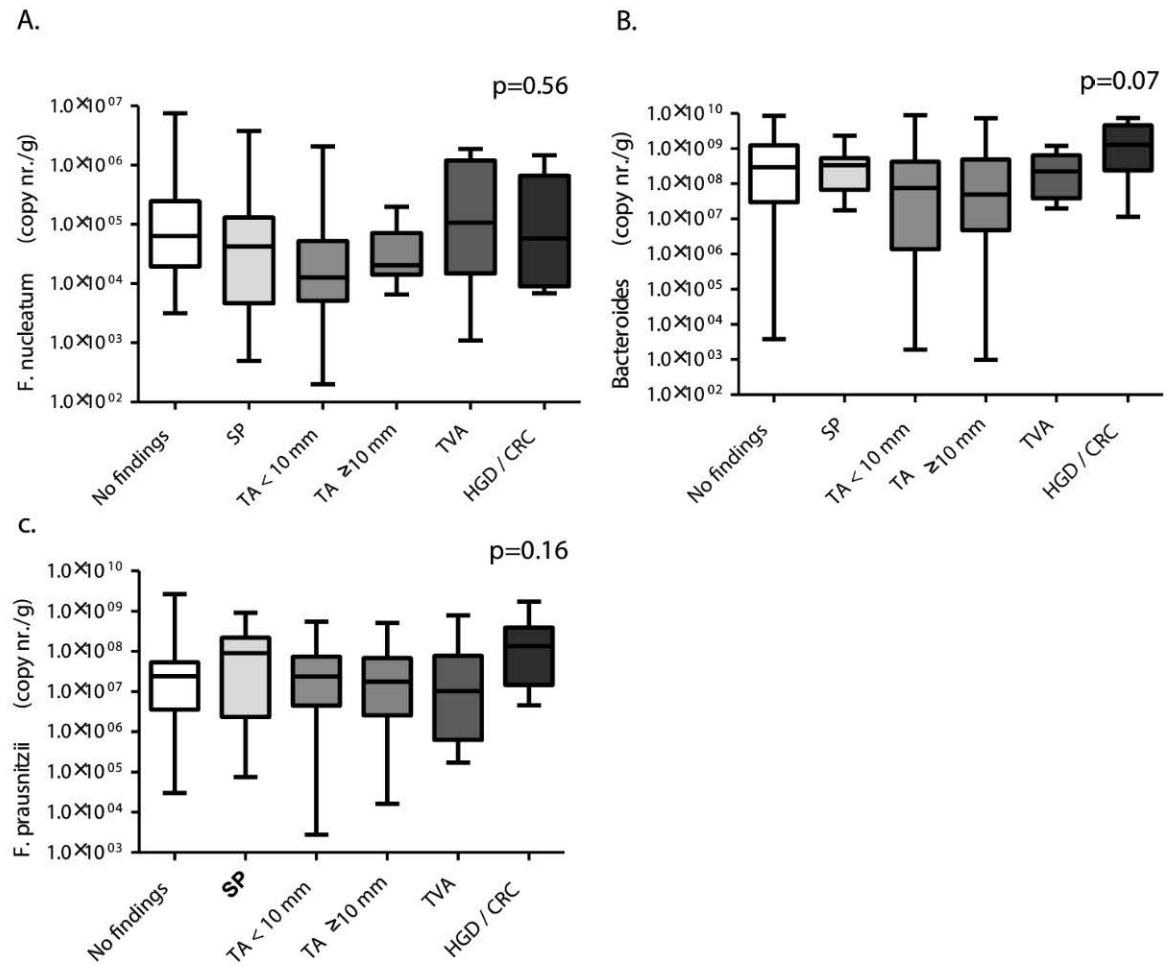
1. Kuipers EJ, Grady WM, Lieberman D, et al. Colorectal cancer. Nature Reviews Disease Primers 2015: article nmber 15065. doi:10.1038/nrdp.2015.65.
2. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. Gut 2016: [E-pub ahead of print] doi:10.1136/gutjnl-2015-310912.
3. Haggar FA, Boushey RP. Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. Clin Colon Rectal Surg 2009; 22(4): 191-7.
4. Cipe G, Idiz UO, Firat D, Bektasoglu H. Relationship between intestinal microbiota and colorectal cancer. World J Gastrointest Oncol 2015; 7(10): 233-40.
5. Narayanan V, Peppelenbosch MP, Konstantinov SR. Human fecal microbiome-based biomarkers for colorectal cancer. Cancer Prev Res (Phila) 2014; 7(11): 1108-11.
6. Gagniere J, Raisch J, Veziant J, et al. Gut microbiota imbalance and colorectal cancer. World J Gastroenterol 2016; 22(2): 501-18.
7. zur Hausen H. The search for infectious causes of human cancers: where and why (Nobel lecture). Angew Chem Int Ed Engl 2009; 48(32): 5798-808.
8. Klein RS, Recco RA, Catalano MT, Edberg SC, Casey JI, Steigbigel NH. Association of Streptococcus bovis with carcinoma of the colon. N Engl J Med 1977; 297(15): 800-2.
9. Kuipers EJ, de Jong A. Gastrointestinal disorders and Streptococcus bovis bacteremia. Ned Tijdschr Geneeskd 1990; 134(28): 1337-9.
10. Sears CL, Garrett WS. Microbes, microbiota, and colon cancer. Cell Host Microbe 2014; 15(3): 317-28.
11. Tjalsma H, Boleij A, Marchesi JR, Dutilh BE. A bacterial driver-passenger model for colorectal cancer: beyond the usual suspects. Nat Rev Microbiol 2012; 10(8): 575-82.
12. Schistosomes, liver flukes and Helicobacter pylori. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Lyon, 7-14 June 1994. IARC Monogr Eval Carcinog Risks Hum 1994; 61: 1-241.
13. Yu YN, Fang JY. Gut Microbiota and Colorectal Cancer. Gastrointest Tumors 2015; 2(1): 26-32.
14. Ito M, Kanno S, Nosho K, et al. Association of Fusobacterium nucleatum with clinical and molecular features in colorectal serrated pathway. Int J Cancer 2015; 137(6): 1258-68.
15. Goedert JJ, Gong Y, Hua X, et al. Fecal Microbiota Characteristics of Patients with Colorectal Adenoma Detected by Screening: A Population-based Study. EBioMedicine 2015; 2(6): 597-603.
16. Flanagan L, Schmid J, Ebert M, et al. Fusobacterium nucleatum associates with stages of colorectal neoplasia development, colorectal cancer and disease outcome. Eur J Clin Microbiol Infect Dis 2014; 33(8): 1381-90.
17. Sears CL, Pardoll DM. Perspective: alpha-bugs, their microbial partners, and the link to colon cancer. J Infect Dis 2011; 203(3): 306-11.
18. Underwood MA. Intestinal dysbiosis: novel mechanisms by which gut microbes trigger and prevent disease. Prev Med 2014; 65: 133-7.
19. Schreuders EH, Ruco A, Rabeneck L, et al. Colorectal cancer screening: a global overview of existing programmes. Gut 2015; 64(10): 1637-49.
20. van Doorn SC, Stegeman I, Stroobants AK, et al. Fecal immunochemical testing results and characteristics of colonic lesions. Endoscopy 2015; 47(11): 1011-7.

21. Lee JK, Liles EG, Bent S, Levin TR, Corley DA. Accuracy of fecal immunochemical tests for colorectal cancer: systematic review and meta-analysis. Ann Intern Med 2014; 160(3): 171.
22. Lieberman D. Screening for colorectal cancer in individuals at average risk: current methods and emerging issues. JAMA Intern Med 2014; 174(1): 10-1.
23. de Wijkerslooth TR, Stoop EM, Bossuyt PM, et al. Immunochemical fecal occult blood testing is equally sensitive for proximal and distal advanced neoplasia. Am J Gastroenterol 2012; 107(10): 1570-8.
24. Lopez-Siles M, Martinez-Medina M, Suris-Valls R, et al. Changes in the Abundance of Faecalibacterium prausnitzii Phylogroups I and II in the Intestinal Mucosa of Inflammatory Bowel Disease and Patients with Colorectal Cancer. Inflamm Bowel Dis 2016; 22(1): 28-41.
25. Kapidzic A, Grobbee EJ, Hol L, et al. Attendance and yield over three rounds of population-based fecal immunochemical test screening. Am J Gastroenterol 2014; 109(8): 1257-64.
26. van Roon AH, Wilschut JA, Hol L, et al. Diagnostic yield improves with collection of 2 samples in fecal immunochemical test screening without affecting attendance. Clin Gastroenterol Hepatol 2011; 9(4): 333-9.
27. Schlemper RJ, Riddell RH, Kato Y, et al. The Vienna classification of gastrointestinal epithelial neoplasia. Gut 2000; 47(2): 251-5.
28. Sokol H, Pigneur B, Watterlot L, et al. Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. Proc Natl Acad Sci U S A 2008; 105(43): 16731-6.
29. Periasamy S, Chalmers NI, Du-Thumm L, Kolenbrander PE. Fusobacterium nucleatum ATCC 10953 requires Actinomyces naeslundii ATCC 43146 for growth on saliva in a three-species community that includes Streptococcus oralis 34. Appl Environ Microbiol 2009; 75(10): 3250-7.
30. Borgo PV, Rodrigues VA, Feitosa AC, Xavier KC, Avila-Campos MJ. Association between periodontal condition and subgingival microbiota in women during pregnancy: a longitudinal study. J Appl Oral Sci 2014; 22(6): 528-33.
31. Huijsdens XW, Linskens RK, Mak M, Meuwissen SG, Vandenbroucke-Grauls CM, Savelkoul PH. Quantification of bacteria adherent to gastrointestinal mucosa by real-time PCR. J Clin Microbiol 2002; 40(12): 4423-7.
32. Carasi P, Racedo SM, Jacquot C, Romanin DE, Serradell MA, Urdaci MC. Impact of kefir derived Lactobacillus kefiri on the mucosal immune response and gut microbiota. J Immunol Res 2015; 2015: 361604.
33. Kabeerdoss J, Ferdous S, Balamurugan R, et al. Development of the gut microbiota in southern Indian infants from birth to 6 months: a molecular analysis. J Nutr Sci 2013; 2: e18.
34. Malinen E, Rinttila T, Kajander K, et al. Analysis of the fecal microbiota of irritable bowel syndrome patients and healthy controls with real-time PCR. Am J Gastroenterol 2005; 100(2): 373-82.
35. Ahmed S, Macfarlane GT, Fite A, McBain AJ, Gilbert P, Macfarlane S. Mucosa-associated bacterial diversity in relation to human terminal ileum and colonic biopsy samples. Appl Environ Microbiol 2007; 73(22): 7435-42.
36. Aguirre de Carcer D, Cuiv PO, Wang T, et al. Numerical ecology validates a biogeographical distribution and gender-based effect on mucosa-associated bacteria along the human colon. ISME J 2011; 5(5): 801-9.
37. Liguori G, Lamas B, Richard ML, et al. Fungal Dysbiosis in Mucosa-associated Microbiota of Crohn's Disease Patients. J Crohns Colitis 2016; 10(3): 296-305.
38. Liguori G, Lamas B, Richard ML, et al. Fungal Dysbiosis in Mucosa-associated Microbiota of Crohn's Disease Patients. J Crohns Colitis 2015.
39. Schwabe RF, Jobin C. The microbiome and cancer. Nat Rev Cancer 2013; 13(11): 800-12.
40. Imperiale TF, Ransohoff DF, Itzkowitz SH, et al. Multitarget stool DNA testing for colorectal-cancer screening. N Engl J Med

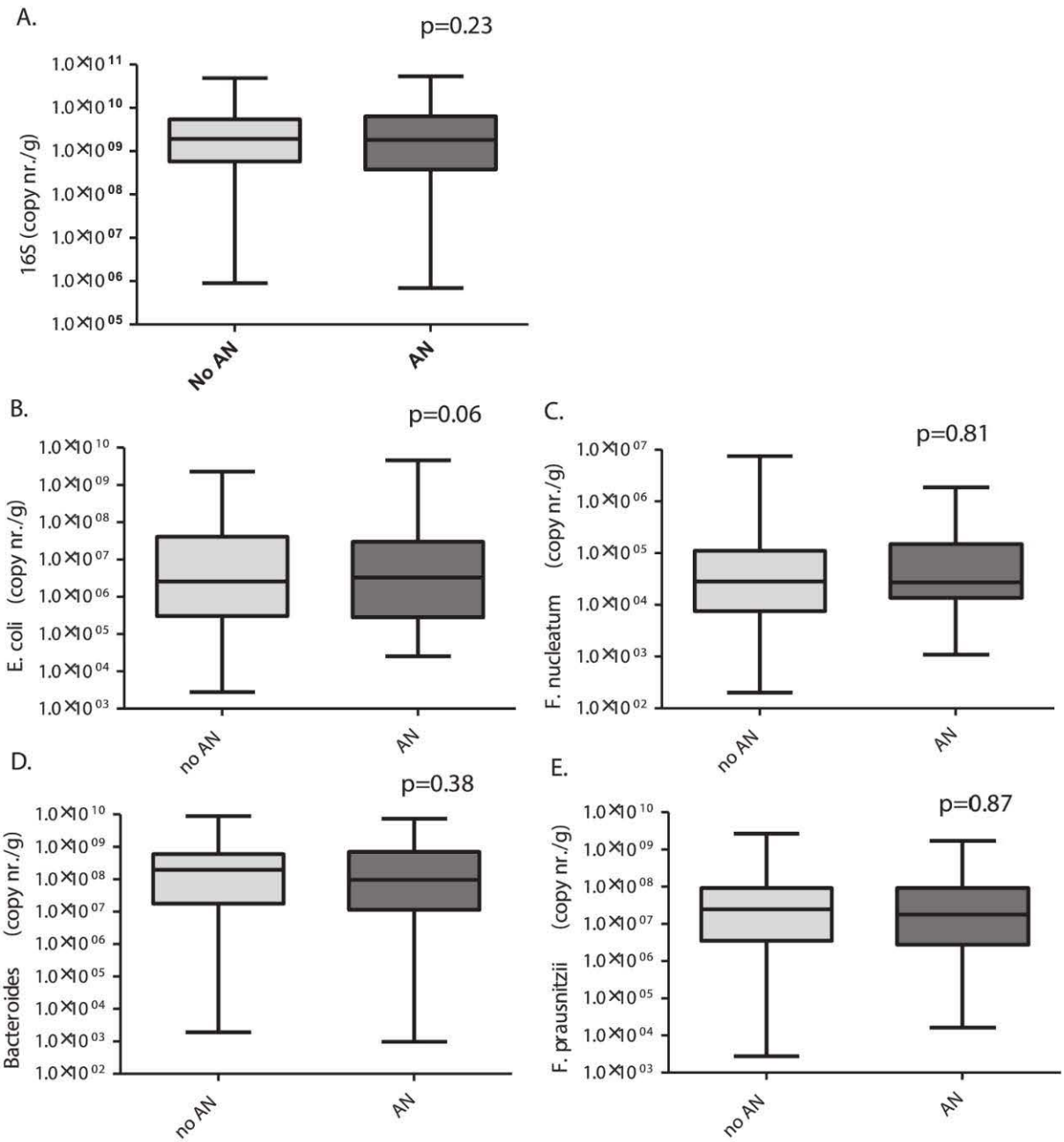
2014; 370(14): 1287-97.

41. Zoetendal EG, von Wright A, Vilpponen-Salmela T, Ben-Amor K, Akkermans AD, de Vos WM. Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. Appl Environ Microbiol 2002; 68(7): 3401-7.
42. Deeks JJ, Bossuyt PM, Gatsonis C. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy, Version 1.0.0. The Cochrane Collaboration; 2009.
43. Baxter NT, Ruffin MTt, Rogers MA, Schloss PD. Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. Genome Med 2016; 8(1): 37.
44. Leeflang MM, Moons KG, Reitsma JB, Zwinderman AH. Bias in sensitivity and specificity caused by data-driven selection of optimal cutoff values: mechanisms, magnitude, and solutions. Clin Chem 2008; 54(4): 729-37.
45. Schroedl W, Kleessen B, Jaekel L, Shehata AA, Krueger M. Influence of the gut microbiota on blood acute-phase proteins. Scand J Immunol 2014; 79(5): 299-304.
46. Stegeman I, van Doorn SC, Mundt MW, et al. Participation, yield, and interval carcinomas in three rounds of biennial FIT-based colorectal cancer screening. Cancer Epidemiol 2015; 39(3): 388-93.

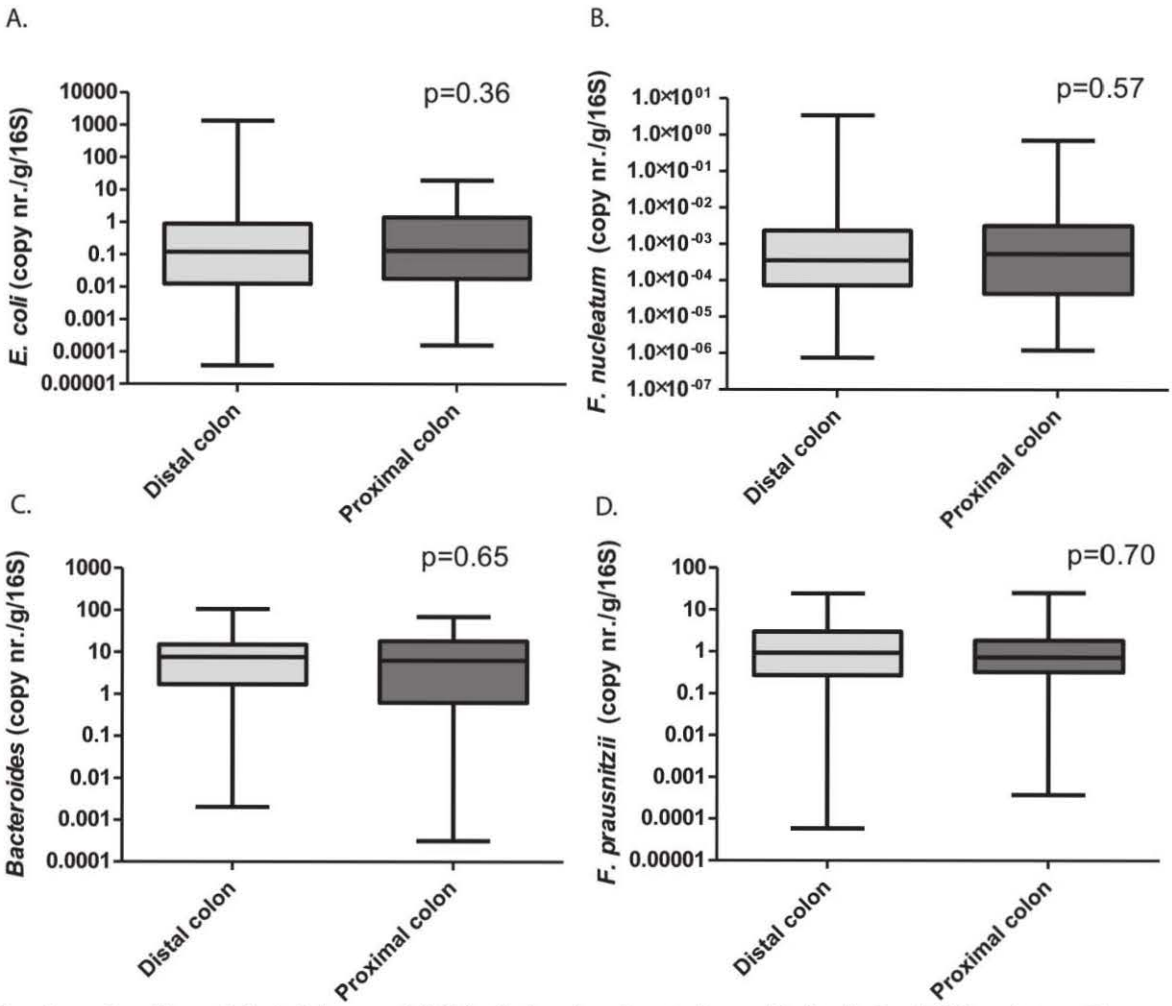
Supplementary files



Supplementary Figure 1. *Bacterial copy nr./g FIT liquid for F. nucleatum (A), Bacteroides (B) and F. prausnitzii.(C) for the most advanced colonoscopy finding*.*



Supplementary Figure 2. *Bacterial copy nr./g FIT liquid for 16S (A), E. Coli (B) F. Nucleatum (C), Bacteroides (D) and F. Prausnitzii (E) for advanced neoplasia.*



Supplementary Figure 3. *Bacterial copy nr./g/16S for the location of most advanced finding for E. coli (A) F. nucleatum (B), Bacteroides (C) and F. prausnitzii (D).*

PART III

using FIT as a quantitative guide in in colorectal cancer screening


chapter
8

Fecal hemoglobin concentrations predict future advanced colorectal neoplasia in long-term population-based FIT-screening

Esmée J. Grobbee, Eline H. Schreuders, Bettina E. Hansen, Marco J. Bruno, Iris Lansdorp-Vogelaar, Manon C.W. Spaander, Ernst J. Kuipers

Submitted

Abstract

Background and aims Colorectal cancer (CRC) screening using quantitative fecal immunochemical tests (FITs) is rapidly gaining ground worldwide. FITs are invariably used in a dichotomous manner using pre-specified cut-offs. To optimize FIT-based screening programs, we explored if fecal hemoglobin concentrations ($_f$Hb) of participants with a FIT below the cut-off (FIT_{bco}) could be used to predict future colorectal advanced neoplasia (AN) risk.

Methods Average-risk subjects aged 50-74 years, were offered four rounds of population-based FIT screening (cut-off 10 µg Hb/g feces). All subjects with a FIT_{bco} at first participation (baseline) were included. Hazard ratios (HRs) for AN were determined using Cox proportional hazard regression analyses. Logistic regression techniques were used to calculate risks of AN after consecutive FIT_{bco} results.

Results Out of 13,566 invitees, 9,561 (70%) participated at least once and 7,663 (92%) had FIT_{bco} at baseline. Median follow up was 4.7 years (IQR 2.0-6.1). After eight years of follow-up, a higher cumulative incidence of AN was found in screenees with baseline $_f$Hb between 8 to 10 µg Hb/g compared to those with $_f$Hb of 0 µg Hb/g (5 vs. 33%; $p<0.001$). The multivariate HRs increased from 1.2 to 8.2 for $_f$Hb concentrations between >0 and 2 µg Hb/g and ≥8 to 10 µg Hb/g ($p<0.001$). A 14-fold increased risk was found after two consecutive $FITs_{bco}$ with twice $_f$Hb of 8 µg Hb/g versus twice 0 µg Hb/g ($p<0.001$).

Conclusion Among screenees with a FIT_{bco}, baseline and consecutive $_f$Hb are independent predictors for incident AN. These findings provide tools for personalized strategies in population-based CRC screening. This may decrease unnecessary screenee burden and could optimize use of resources.

Introduction

Colorectal cancer (CRC) is among the most common causes of cancer-related mortality[1]. Population-based CRC screening can significantly reduce disease burden. Fecal occult blood tests are widely accepted for this purpose[2,3].A higher fecal hemoglobin ($_f$Hb) concentration is associated with a higher risk of advanced neoplasia (AN)[4-7].Many screening programs worldwide use fecal immunochemical tests (FIT), that can be either qualitative (i.e. providing a positive or negative test result) or quantitative (i.e. quantifying $_f$Hb concentrations in feces)[2,8]. Although quantitative FITs provide exact fHb concentrations in µg Hb/g feces, current screening programs routinely use $_f$Hb in a dichotomized fashion. As such, they are invariably used as qualitative tests. A test is considered positive above a fixed threshold that is the same for all screenees in all rounds of screening. Those with a positive test are recommended to undergo colonoscopy. Individuals with a negative test are offered a renewed FIT after a predefined screening interval without taking into account previous fHb concentrations. Most FIT screening programs rely on annual or biennial screening, requiring participants to repeat the test multiple times over the course of years.

To increase screening efficiency and impact of FIT screening programmes, it is relevant to explore if screenees with a negative FIT, that is a $_f$Hb concentration below the pre-defined cut-off level (FIT below cut-off; FIT_{bco}), can be categorized according to their actual $_f$Hb concentration into different risk groups for later development of AN. Such tailored screening would allow for targeted variation of screening intervals, and decrease screening and colonoscopy demand or optimize its use. Currently, many countries with CRC screening programs struggle to match colonoscopy demands with limited resources[9-11]. Over the course of multiple screening rounds, $_f$Hb concentration could then be of guidance in identifying those at low and high risk of AN, and thus form the basis of individualized screening strategies. Such information is of key importance for national population-based CRC screening policies. Previous studies have mainly focused on $_f$Hb concentrations of FIT-positive screenees, or have only assessed first round $_f$Hb concentrations[5,12,13]. At present, no literature is available on trends in individual $_f$Hb concentrations of FIT_{bco} screenees over consecutive screening rounds. Furthermore, it is not known whether these previous $_f$Hb concentration(s) can be used as a predictor for the future detection of AN. Therefore, we aimed to investigate trends in $_f$Hb of FIT_{bco} results at first participation, and in subsequent rounds as a predictor for future incidence of AN.

Methods

Study design and participants

Details about this study cohort and the design have been described before[14]. In short, individuals living in the southwest of the Netherlands were approached for four rounds of FIT-screening. Demographic data of all individuals between 50 and 74 years living in this region were obtained from municipal population registers. Random samples were taken based on different postal codes. Exclusion criteria included a history of inflammatory bowel disease or CRC, a full colonic examination in the past two years, an estimated life expectancy of <5 years, and inability to give informed consent. In case of a positive FIT result, subjects were sent for colonoscopy and not re-invited for subsequent FIT rounds. Subjects were not invited when they had become older than 74 years, or when they had moved out of the region. The cohort was supplemented with new screening-naïve subjects in round 3 and 4 to best mimic a continuous, population-based screening cohort. Recruitment took place between November 2006 and December 2014. For this analysis all subjects participating in at least one screening round and with a negative FIT result, defined as a FIT result below the cut-off of 10 µg Hb/ g feces (FIT_{bco}) at their first participation (i.e. baseline), were included.

FIT screening and colonoscopy

Each screening round, eligible invitees received one FIT per mail. Invitees were instructed to collect one sample of one bowel movement. In the first 3 rounds subjects received the OC-sensor (Eiken, Japan), and in the last round subjects were randomized to receive either the OC-sensor or the FOB-Gold (Sentinel, Italy). The OC-Sensor and FOB-Gold perform equally over the relevant concentration range[15]. Participants were asked to sample feces according to instructions and post the sample together with the consent form within 24 hours while storing it in the refrigerator until mailing. The OC sensor FITs were analysed on the OC-sensor µ system (Eiken, Japan), the FOB-gold FITs were analysed on a Sentifit 270 system (Sentinel, Italy). All FIT tests were analysed once at room temperature. The analytical working ranges for the OC sensor µ and Sentifit 270 were respectively 1-200 feces and 1-170 µg Hb/g feces. Samples with fHb concentrations above the upper analytical working limits were not diluted or re-analysed. The test result of ≥10 µg Hb/g feces was considered positive. Subjects with a positive FIT were scheduled for colonoscopy within 4 weeks. In case colonoscopy was incomplete, a computed tomographic colonography was performed.

Experienced board-certified gastroenterologists performed all endoscopies. The maximum reach of the endoscope, adequacy of bowel preparation, and characteristics and location of any polyps

were recorded. All polyps were removed and evaluated by dedicated gastrointestinal pathologists. Patients with a positive colonoscopy entered a surveillance program according to guidelines of the Dutch Society of Gastroenterology, whereas subjects with a negative colonoscopy were considered not to require FIT screening for 10 years. AN was defined as an adenoma of ≥ 10 mm, an adenoma with at least 25% villous histology and/or high-grade dysplasia, or CRC.

Follow-up and interval carcinomas

Except for individuals who had moved out of the Netherlands, all recruited participants were followed for the development of CRC. Colorectal cancers diagnosed outside of the screening program (including CRCs in participants who did not return to screening, FIT interval cancers, post-colonoscopy CRCs, and CRCs detected at surveillance colonoscopy) were identified through linkage with the Dutch Comprehensive Cancer Centre (www.iknl.nl), which was up to date until March 2015.

Statistical analysis

Descriptive data were reported as proportions or means with standard deviation (±SD). For non-normally distributed data, the median and interquartile range (IQR) were given. To investigate the role of a negative baseline FIT value, $_{f}$Hb concentrations were divided into six categories; 0, >0-2, ≥2-4, ≥4-6, ≥6-8 and ≥8-10. Per category the cumulative incidence of AN over four rounds was calculated using life tables and curves. Patients were censored at the end of follow up if the event (i.e. AN) had not occurred. A Cox proportional hazard regression analysis was performed to calculate hazard ratios (HRs), including 95% confidence intervals, to identify factors associated with the development of AN. These factors included age, gender, socioeconomic status (SES), and baseline $_{f}$Hb. The date of the baseline FIT_{bco} was defined as time 0. Only $_{f}$Hb concentrations below 10 µg Hb/g feces were used in these analyses. Factors with a p-value of < 0.10 in univariate analysis, were included in a multivariate model. Interaction terms were also evaluated in the multivariate model. A two-sided P value of < 0.05 was considered to be statistically significant. Analyses were performed using SPSS 21.0 statistics software (IBM Corp., Armonk, New York, USA).

The prediction of an event (i.e. AN or CRC) given the outcome of $_{f}$Hb concentration of a FIT_{bco} at any round (i.e. visit) was analysed applying a logistic regression technique allowing for multiple repeated measurements per subject using SAS PROC GENMOD with the REPEATED statement with an independent variance assumption[16]. Each visit was hereby associated with the event. The analysis included adjustment for age at first round (or age at time of FIT screening), sex, and time of round in relation to event or last follow-up. Interactions between gender and age as well as non-

linearity of age and $_{f}$Hb concentration were tested using both pre-specified groups and polynomial regression. Factors with a p-value < 0.10 in univariate analysis, were included in the multivariate model. A two-sided p-value < 0.05 was considered to be statistically significant. In addition, all analyses were adjusted for multiple testing. Using the results of logistic regression analyses heat plots were generated to depict the risk of AN after two FIT_{bco}. Analyses were performed using SAS 9.4 (SAS Institute Inc., Cary, North Carolina, USA).

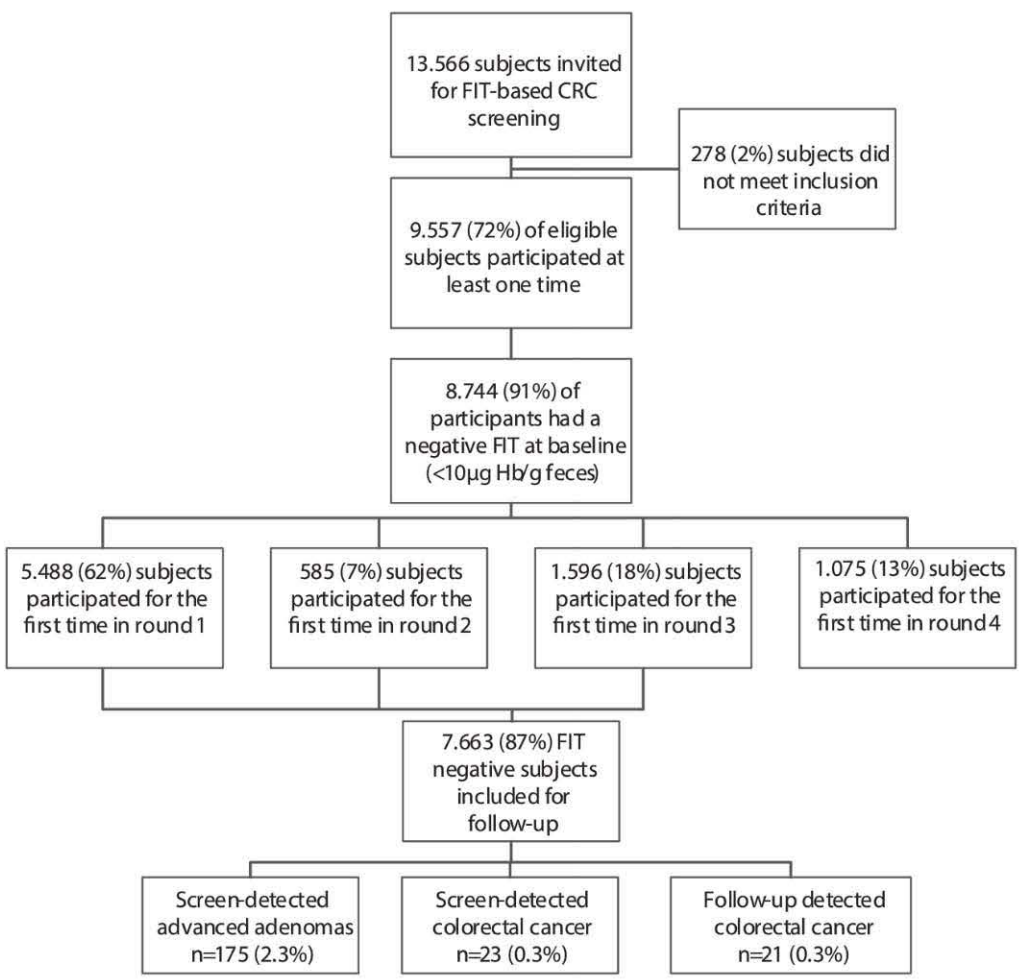


Figure 1. *Flowchart of study design and carcinomas.*

Results

Baseline characteristics

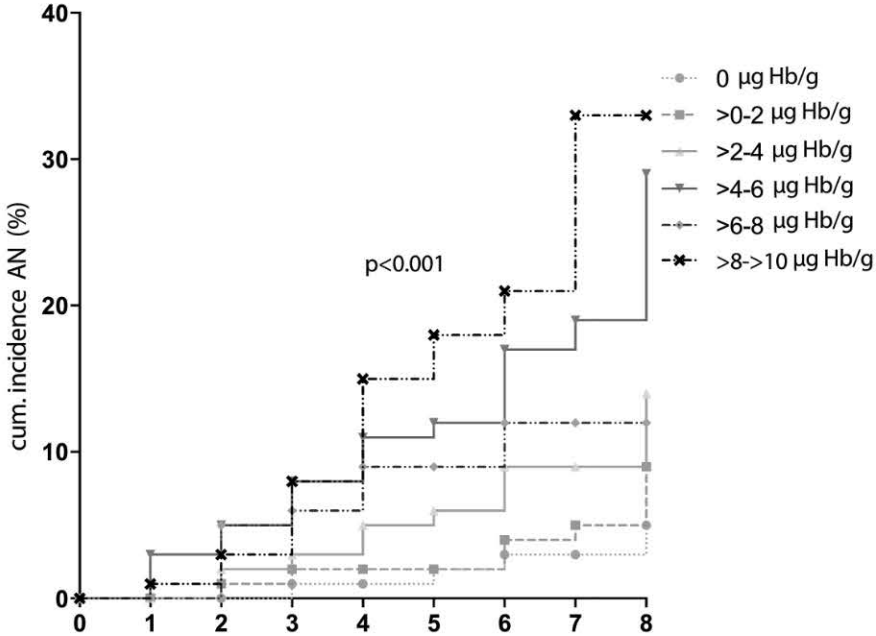
Over four rounds of biennial FIT screening, a total 13,566 subjects were invited of whom 278 (2.0%) did not meet the inclusion criteria. Out of the 13,288 remaining subjects 9,557 (71.9%) participated at least once (Figure 1). Out of these participants, 8,744 (91.4%) had a FIT_{bco} at first participation and were included for analysis. Median age was 58 years (IQR 52-64 years) at baseline, with 47·1% of the screenees being male. Overall, 3,172 (36.3%) FIT_{bco} subjects participated in all four rounds, 1,235 (14.3%) in three rounds, 2,254 (25.8%) in two rounds, and 2·059 (23.6%) in one round. Median time between screening rounds was 23 months (IQR 22 - 24 months).

Role of baseline $_f$Hb in predicting risk of advanced neoplasia and colorectal caner

The majority of subjects (62.7%) participated for the first time in round one, the remaining subjects participated in round two (6.6%), round three (18.3%), and round four (12.3%). Because follow-up ended after the fourth round, only baseline $_f$Hb of subjects participating in one of the first three rounds (n=7,663) were included for survival analyses. Median follow up was 4.7 years (IQR 2.0-6.1 years). Over the following rounds, 821 (10.7%) participants with a baseline FIT_{bco} had a positive FIT, of whom 91.8% underwent colonoscopy (n=754). During follow-up 221 (3.0%) cases were diagnosed with AN; 175 with advanced adenomas, 24 with screen-detected CRC, and 22 with CRC detected during follow-up, of which 8 were FIT interval cancers (Figure 1).

Cumulative incidences of AN per category of $_f$Hb are shown in Figure 2. After 5 years of follow-up, screenees with a baseline $_f$Hb concentration between 8 and 10 µg Hb/g had an 8-fold higher cumulative incidence of AN than screenees with a baseline $_f$Hb concentration of 0 µg Hb/g ($p < 0·001$). After 8 years of follow-up the cumulative incidence increased with 28% compared to screenees with a baseline $_f$Hb concentration of 0 µg Hb/g feces ($p<0.001$).

Cox proportional hazard regression analysis showed baseline $_f$Hb concentration was associated with the hazard of developing AN in multivariate analysis (Table 1). Compared to screenees with a baseline $_f$Hb concentration of 0 µg Hb/g feces, HRs increased from 1.2 (95% CI 0.9-1.7) for fHb concentrations between >0 and 2 µg Hb/g, to 8.2 (95% CI 4.5-15.0) for $_f$Hb concentrations of ≥8 to 10 µg Hb/g ($p<0.001$).



fHb (µg Hb/g)	1	2	3	4	5	6	7	8	Time (years)
0	4,927	4,185	3,639	2,852	2,410	2,090	1,485	726	Subjects at risk
>0-2	1,874	1,672	1,587	1,427	1,291	1,177	470	94	Subjects at risk
≥2-4	436	376	333	286	247	220	101	34	Subjects at risk
≥4-6	214	171	151	125	105	94	54	16	Subjects at risk
≥6-8	106	87	78	65	56	47	18	4	Subjects at risk
≥8-10	78	70	58	44	35	29	18	7	Subjects at risk

Figure 2. *Life table and curve for advanced neoplasia by $_f$Hb level per 2µg Hb/g. This figure shows that the effect on cumulative incidence of AN of baseline FIT is most prominent for fHb between 4 and 10 µg Hb/g.*

Role of $_f$Hb concentrations of consecutive FITs in predicting risk of advanced neoplasia

To further explore the use of $_f$Hb of consecutive FITs, $_f$Hb concentrations of all screenees that had at least two consecutive $FITs_{bco}$ were analysed. In multivariate logistic regression, including time between FITs and total follow-up time, several FIT_{bco} combinations were evaluated to determine relative risk of developing AN, these are shown in Table 2. ORs remained similar regardless of the sequence of $_f$Hb results.

Table 1. *Cox-regression analysis of baseline $_f$Hb concentration for hazard of the detection of advanced neoplasia during follow-up.*

	Advanced neoplasia			
	Univariate analysis		**Multivariate analysis**	
	HR	95% CI	HR	95%CI
Gender (male)	1·7*	1.3-2·3	1·6*	1·2-2·1
Age (years)	1·1*	1·0-1·1	1·1*	1·0-1·1
Baseline $_f$Hb concentration				
0 µg Hb/g	Ref.*		Ref.*	
>0-2 µg Hb/g	1·3	0·9-1·8	1·2	0·9-1·7
≥2-4 µg Hb/g	2·9	1·8-4·6	2·8	1·7-4·4
≥4-6 µg Hb/g	6·5	4·2-10·2	5·7	3·7-8·9
≥6-8 µg Hb/g	4·5	2·1-9·3	4·2	2·1-8·7
≥8-10 µg Hb/g	8·9	4·9-16·4	8·2	4·5-15·0
Socioeconomic status				
High	Ref.			
Average	1·0	0·7-1·3		
Low	0·6	0·4-1·0		

* p<0·001

$_f$Hb; fecal hemoglobin concentration, HR; hazard ratio, CI; confidence interval, conc; concentration, ref; reference category.

Table 2. *Multivariate logistic regression analysis evaluating various hypothetical combinations of fHb concentration of two previously negative FITs and odds ratio (OR) of developing advanced neoplasia during follow-up.*

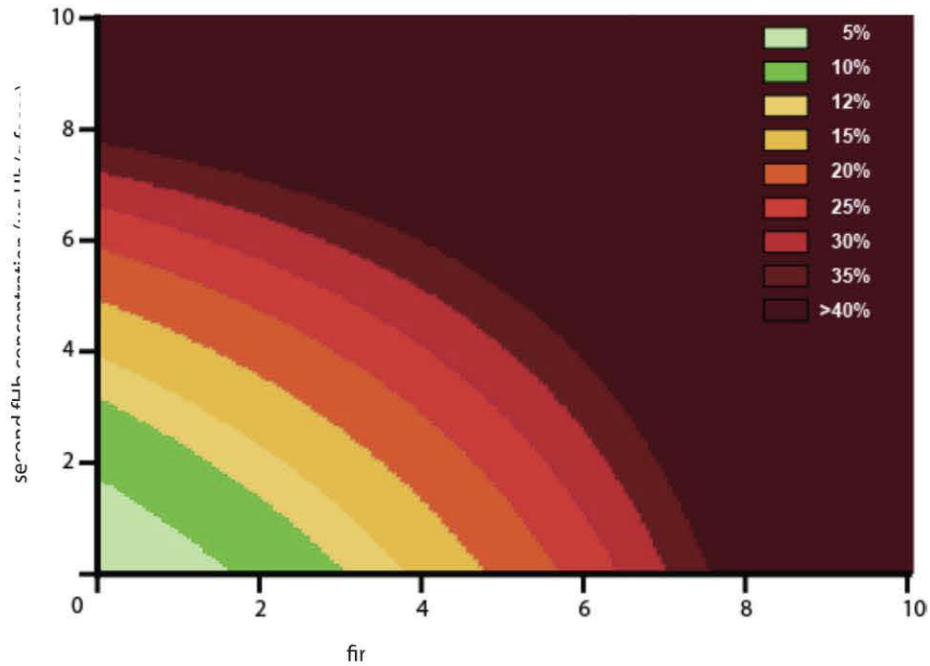
	Advanced neoplasia		
	Multivariate analysis		
	OR	95% CI	p-value
Gender (female)	2·1	1·3-3·2	0·001
Age (years)	1·0	1·0-1·1	0·04
Combination of first and second fHb concentration			
0 µg Hb/g and 0 µg Hb/g	Ref.		<0·001
1 µg Hb/g and 1 µg Hb/g	1·7	1·5-1·9	
1 µg Hb/g and 5 µg Hb/g	4·4	3·1-6·3	
5 µg Hb/g and 1 µg Hb/g	4·5	3·1-6·6	
5 µg Hb/g and 5 µg Hb/g	7·8	4·6-13·3	
1 µg Hb/g and 8 µg Hb/g	9·0	5·2-15·6	
8 µg Hb/g and 8 µg Hb/g	14·3	4·8-42·3	

fHb; fecal hemoglobin concentration, HR; hazard ratio, CI; confidence interval, conc; concentration, ref; reference category

Based on the logistic regression model risks were predicted for all possible combinations of $_f$Hb concentrations per 0.5 µg Hb/g feces. Using these risks, heat plots were generated for males and females starting screening from the age of 55 years (Figure 3). These heat plots visualize the risk of AN after two consecutive FIT_{bco} at any round according to the combination of $_f$Hb concentrations.

These heat plots highlight the increased risk of screenees with $_{f}$Hb concentrations up to 10 µg Hb/g and illustrate two-fold increased risk of AN for men compared to women.

A.



B.

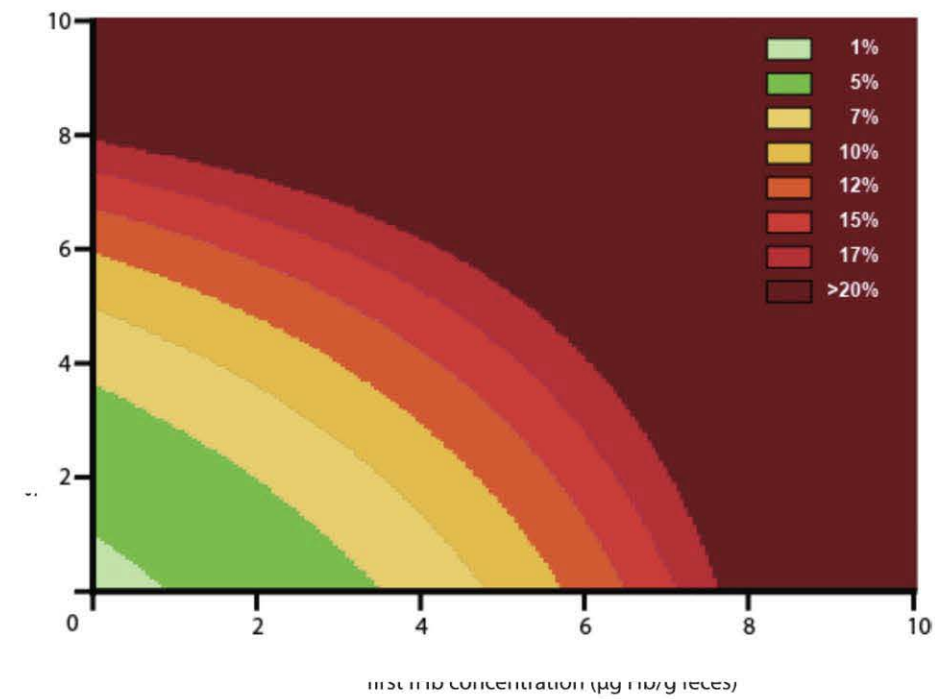


Figure 3. *Heat plot of risk of advanced neoplasia (AN) during further follow up in screenees with two consecutive negative FITs for men (**A**) and women (**B**) starting screening at age 55 years. This plot illustrates the increased risk of AN according to fHb concentration. Notable is the two-fold increase for men compared to women.*

Discussion

This study shows that, in a repeated-round FIT-based CRC screening program, fecal hemoglobin concentration below the cut-off is an independent predictor of incident AN. We demonstrated over 8 years of follow-up that screenees with a $_{f}$Hb concentration >0 µg Hb/g to 10 µg Hb/g have a significantly higher cumulative incidence of AN compared to those with a $_{f}$Hb concentration of 0 µg Hb/g. The risk of AN was 8-fold higher for screenees with a $_{f}$Hb concentration of more than 8 µg Hb/g compared to screenees with a $_{f}$Hb concentration of 0 µg Hb/g. We further show that this predictive capacity becomes even stronger with consecutive FIT_{bco} results obtained over repeated rounds.

One other research group has studied the role of FIT concentrations below the cut-off[12]. This Taiwanese study used a cut-off of 20 µg Hb/g feces (OC-Sensor 100 ng Hb/ml buffer), and showed that HRs for AN during follow-up increased with $_{f}$Hb concentration up to 3.4 for subjects with an $_{f}$Hb level of 16-20 µg Hb/g feces. This HR is substantially lower than the HRs found in our study. This can be explained by the fact that the investigators used $_{f}$Hb concentrations between 1-4 µg Hb/g feces as a reference and screenees with a $_{f}$Hb concentration of 0 were not included in multivariate survival analysis. The results of this study were however hampered by the high rate of positive FIT screenees who did not undergo colonoscopy (42%) yet were included in the analyses. While there is no other literature on the use of quantitative FIT_{bco} results, a few studies did look at prior FIT_{bco} results of qualitative tests. An Australian study investigated the use of FIT in a colonoscopy surveillance program and found that subjects with a FIT_{bco} had the lowest risk of AN[17]. A Chinese study compared the number of FIT_{bco} and found no differences in outcome between subjects with one FIT_{bco} versus subjects with three subsequent FIT_{bco}[18]. However, as the fHb concentrations were not reported, results could not be stratified according to $_{f}$Hb concentration, and comparison of these results to our findings is not possible.

To the best of our knowledge, we are the first to explore the use of $_{f}$Hb concentrations of FIT_{bco} over consecutive rounds as a predictive variable for AN in population-based CRC screening. Exploring $_{f}$Hb concentrations over the course of years makes sense, as it has been hypothesized that during the development from adenoma to carcinoma, in the adenoma-carcinoma sequence, adenomas will increasingly bleed. This natural history of adenomas is supported by our findings. Our results are further strengthened by the finding that in screenees who participated in all four rounds $_{f}$Hb increased among those that were diagnosed with AN in the fourth round (supplementary file 1). A

similar trend was also described by the Taiwanese study, demonstrating that median $_{f}$Hb increases over rounds among screenees that are diagnosed with AN in a later round[12].

Strengths of this study include the analysis over multiple rounds, stratifying for FIT_{bco} levels unto nil µg Hb/g feces. Also, only average risk individuals were included, and the program consisted of true population-based screening, with a consistent screening protocol. This makes these results applicable to all fecal occult blood screening programs worldwide[8]. Our study also has its limitations, such as the limited number of subjects diagnosed with CRC after having participated in three or four rounds. One further issue of this study is that it is susceptible to verification bias, due to the fact that only FIT positive screenees are referred for colonoscopy. As such, the yield of AN could be equally high in screenees with low FIT-values, but these simply do not receive verification by colonoscopy. To partially assess the possibility of verification bias, we performed two additional analysis in which we compared yield of AN only in screenees with a positive FIT during follow-up or screenees with consistent FIT_{bco} (i.e. interval cancer rate; supplementary file 2). Although numbers are small, both analyses consistently showed that the yield of AN was higher in those with higher levels of $_{f}$Hb, similar as our base case analysis. Next, data from our primary colonoscopy screening trail in the same region and time period also showed that FIT_{bco} predicts AN in a single round of FIT (supplementary file 3)[19]. These findings suggest that the impact of verification bias is small and corroborate our finding that FIT_{bco} is predictive for future AN. Despite the verification bias, the lack of colonoscopy in FIT negative screenees, is also a strength of our study. If all screenees would undergo colonoscopy, the opportunity to follow FITbco screenees for the development of AN would be lost.

As more screening programs are being implemented worldwide and FIT is gaining popularity, the use of quantitative FIT should be further explored. Expressing FIT-results not solely as a positive or negative result, but incorporating fHb concentrations in risk prediction models to estimate risk of AN can improve screening efficiency. Our results justify evaluation of screening strategies in which fHb concentrations are used to establish screening intervals. In current practice, a screenee with a FITbco is re-invited in the next screening round to perform a 'new' FIT. This FIT result is used as a referral criterion for colonoscopy regardless of the fHb concentration measured in the previous round. By neglecting the previous FIT result, an opportunity is lost to use the quantitative information of two FITs for risk stratification. We demonstrate, as depicted by the heat plots, that previous FITs enable identification of those at low risk (e.g. screenees who twice had a result of 0 µg Hb/g) and those at considerable risk. Such heat plots can be of use to visualize risks of AN for screenees and health care professionals. Identifying those at high risk of AN would possibly decrease

interval carcinoma rate. Next, combining $_{f}$Hb with known risk factors could facilitate establishing individual screening intervals. Such individualized screening intervals may increase adherence to screening, as the majority of participants (i.e. those with consecutive $_{f}$Hb concentrations of 0 µg Hb/g feces and a low risk of AN) will then have to be screened less frequently. This in turn could increase screening efficiency. As a consequence, screening and subsequently colonoscopy demand would decrease. The latter is especially important, as many countries are struggling with limited colonoscopy capacity[8,9,23].

The results of this prospective FIT-based CRC screening cohort show that $_{f}$Hb concentration of a FIT$_{bco}$ in a first round of screening is an independent predictor for the risk of incident AN. Furthermore, consecutive FIT$_{bco}$ could be used in determining personalized screening strategies. These findings pave the way for the use of $_{f}$Hb concentration in both public health programs as well as clinical decision-making. These results aid in informing patients about the risk of AN after multiple FIT$_{bco}$, and to alter screening intervals accordingly.

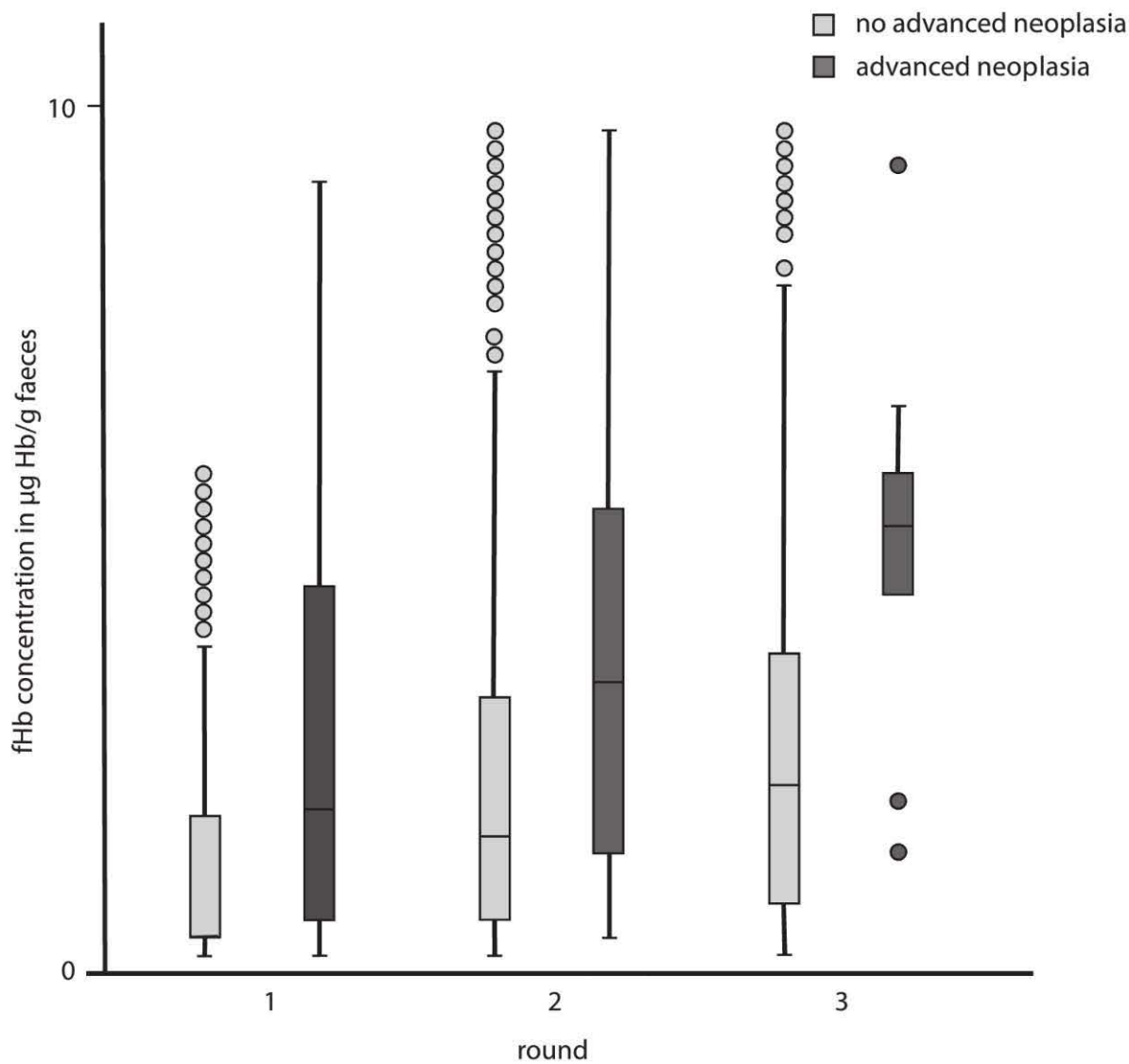
References

1. Kuipers EJ, Grady WM, Lieberman D, et al. Colorectal cancer. Nature Reviews Disease Primers 2015; 1([E-pub ahead of print] 10.1038/nrdp.2015.65):
2. 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.
3. Kuipers EJ, Rosch T, Bretthauer M. Colorectal cancer screening--optimizing current strategies and new directions. Nat Rev Clin Oncol 2013; 10(3): 130-42.
4. van Doorn SC, Stegeman I, Stroobants AK, et al. Fecal immunochemical testing results and characteristics of colonic lesions. Endoscopy 2015; 47(11): 1011-7.
5. Digby J, Fraser CG, Carey FA, et al. Faecal haemoglobin concentration is related to severity of colorectal neoplasia. J Clin Pathol 2013; 66(5): 415-9.
6. Fraser CG, Mathew CM, McKay K,et al. Automated immunochemical quantitation of haemoglobin in faeces collected on cards for screening for colorectal cancer. Gut 2008; 57(9): 1256-60.
7. Hol L, Wilschut JA, van Ballegooijen M, et al. Screening for colorectal cancer: random comparison of guaiac and immunochemical faecal occult blood testing at different cut-off levels. Br J Cancer 2009; 100(7): 1103-10.
8. Schreuders EH, Ruco A, Rabeneck L, et al. Colorectal cancer screening: a global overview of existing programmes. Gut 2015; 64(10): 1637-49.
9. Kanavos P, Schurer W. The dynamics of colorectal cancer management in 17 countries. Eur J Health Econ 2010; 10 Suppl 1: S115-29.
10. Steele RJ, McDonald PJ, Digby J, et al. Clinical outcomes using a faecal immunochemical test for haemoglobin as a first-line test in a national programme constrained by colonoscopy capacity. United European Gastroenterol J 2013; 1(3): 198-205.
11. van Hees F, Zauber AG, van Veldhuizen H, et al. The value of models in informing resource allocation in colorectal cancer screening: the case of The Netherlands. Gut 2015; 64(12): 1985-97.
12. Chen LS, Yen AM, Chiu SY, et al. Baseline faecal occult blood concentration as a predictor of incident colorectal neoplasia: longitudinal follow-up of a Taiwanese population-based colorectal cancer screening cohort. Lancet Oncol 2011; 12(6): 551-8.
13. Ciatto S, Martinelli F, Castiglione G, et al. Association of FOBT-assessed faecal Hb content with colonic lesions detected in the Florence screening programme. Br J Cancer 2007; 96(2): 218-21.
14. Kapidzic A, Grobbee EJ, Hol L, et al. Attendance and yield over three rounds of population-based fecal immunochemical test screening. Am J Gastroenterol 2014; 109(8): 1257-64.
15. Grobbee EJ, van der Vlugt M, van Vuuren A, et al. Comparison of OC-Sensor and FOB-Gold in Population-Based Colorectal Cancer Screening Based on FIT. Gastroenterology; 148(4): S-160.
16. Hernan MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology 2000; 11(5): 561-70.
17. Lane JM, Chow E, Young GP, et al. Interval fecal immunochemical testing in a colonoscopic surveillance program speeds detection of colorectal neoplasia. Gastroenterology 2010; 139(6): 1918-26.
18. Wong MC, Ching JY, Chan VC, et al. Should prior FIT results be incorporated as an additional variable to estimate risk of colorectal neoplasia? A prospective study of 5,813 screening colonoscopies. PLoS One 2014; 9(12): e114332.
19. de Wijkerslooth TR, de Haan MC, Stoop EM, et al. Study protocol: population screening for colorectal cancer by colonoscopy

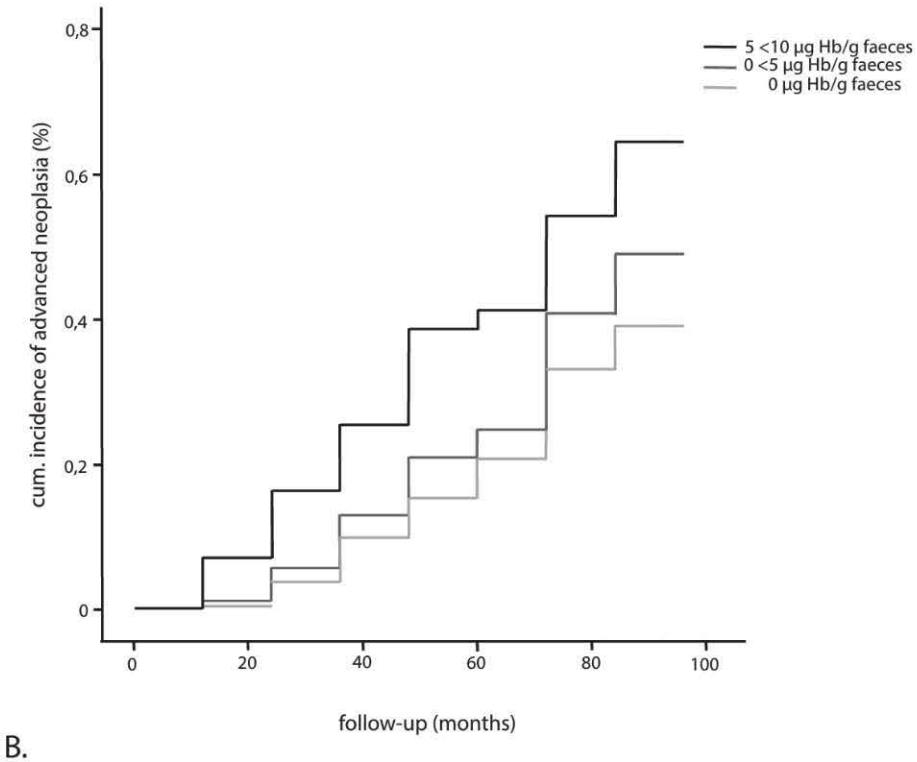
or CT colonography: a randomized controlled trial. BMC Gastroenterol 2010; 10: 47.

20. Quintero E, Castells A, Bujanda L, et al. Colonoscopy versus fecal immunochemical testing in colorectal-cancer screening. N Engl J Med 2012; 366(8): 697-706.
21. Bretthauer M, Kaminski MF, Loberg M, et al. Population-Based Colonoscopy Screening for Colorectal Cancer: A Randomized Clinical Trial. JAMA Intern Med 2016.
22. Brenner H, Chang-Claude J, Jansen L, et al. Colorectal cancers occurring after colonoscopy with polyp detection: sites of polyps and sites of cancers. Int J Cancer 2013; 133(7): 1672-9.
23. Digby J, Fraser CG, Carey FA, et al. Interval cancers using a quantitative faecal immunochemical test (FIT) for haemoglobin when colonoscopy capacity is limited. J Med Screen 2015.

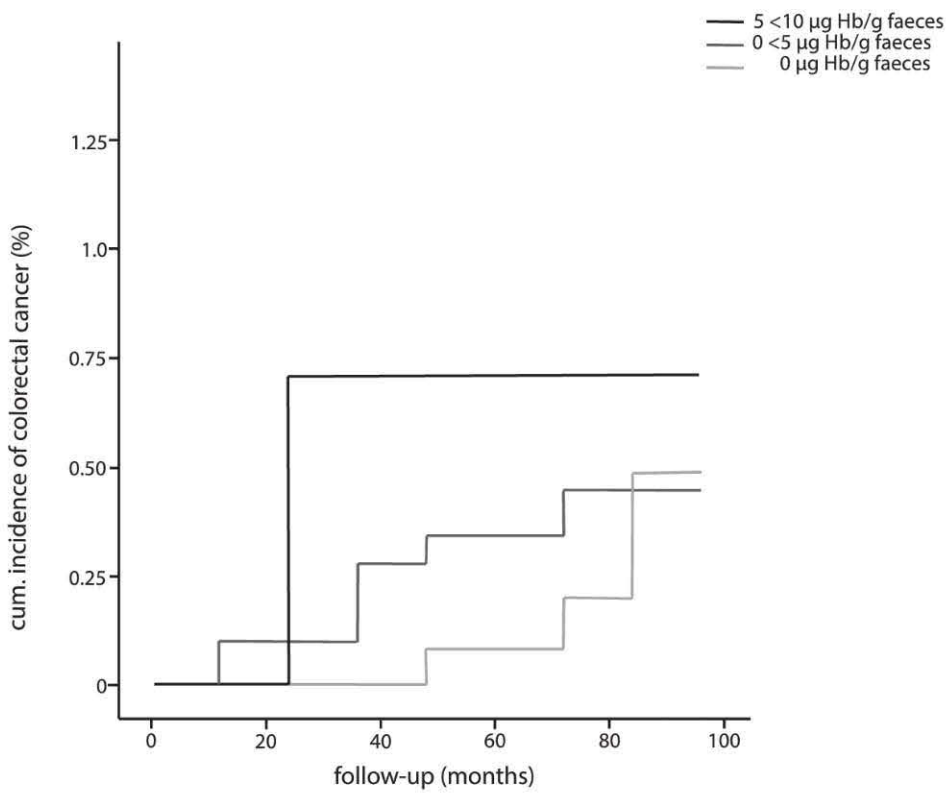
Supplementary files



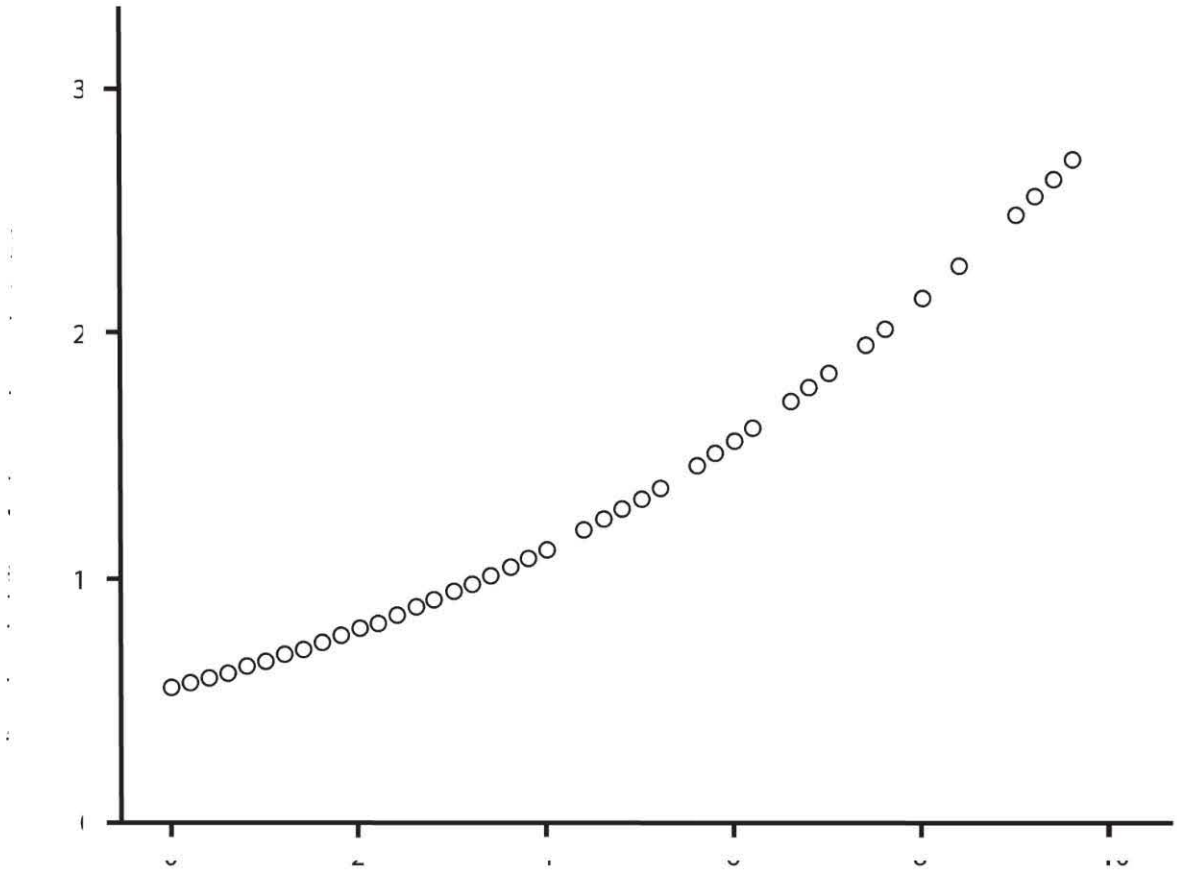
Supplementary Figure 1. *Median fecal Hb of subjects with FIT_{bco} of >0µg Hb/g that participated in all 4 rounds. Results are presented by occurrence of advanced neoplasia (AN) detected in the 4th round. This figure shows an increase in median $_f$Hb concentration over consecutive rounds in subjects who had AN detected in the fourth round, indicating that those screenees that bleed for other reasons than AN, show less of an increase in $_f$Hb over rounds.*



B.



Supplementary Figure 2. *Life table of FIT_{bco} per 5 µg Hb/g in those who had a positive FIT during follow-up (p<0.001; A) and those who never had a positive FIT in one of the four rounds (p=0.034;B).*



Supplementary Figure 3. *Predicted probability of advanced neoplasia (AN) of all subjects with fHb <10 µg Hb/g from previously published data*[19].


chapter
9

Fecal immunochemical test-based colorectal cancer screening: the gender dilemma

Esmée J. Grobbee, Els Wieten, Bettina E. Hansen, Esther M. Stoop, Thomas R. de Wijkerslooth, Iris Lansdorp-Vogelaar, Patrick M. Bossuyt, Evelien Dekker, Ernst J. Kuipers, Manon C.W. Spaander

United European Gastroenterology Journal, 2016

Abstract

Background Despite differences between men and women in incidence of colorectal cancer (CRC) and its precursors, screening programs consistently use the same strategy for both genders.

Objective To illustrate the effects of gender tailored screening, including the effects on miss rates of advanced neoplasia (AN).

Methods Participants (50-75 years) in a colonoscopy screening program were asked to complete a fecal immunochemical test (FIT) before colonoscopy. Positivity rates, sensitivity and specificity for detection of AN at multiple cut-offs were determined. Absolute numbers of detected and missed AN per 1000 screenees were calculated.

Results In total 1,256 underwent FIT and colonoscopy, 51% male (median age 61 years; IQR 56-66) and 49% female (median age 60 years; IQR 55-65). At all cut-offs men had higher positivity rates than women, ranging from 3.8-10.8% versus 3.2-4.8%. Sensitivity for AN was higher in men than women; 40-25% and 35-22%, respectively. More AN were found and missed in absolute numbers in men at all cut-offs.

Conclusion More AN were both detected and missed in men compared to women at all cut-offs. Gender tailored cut-offs could either level sensitivity in men and women (i.e. lower cut-off in women) or level the amount of missed lesions (i.e. lower cut-off in men).

Introduction

Colorectal cancer (CRC) is one of the leading causes of cancer-related death in the Western world.1,2 Detection of occult blood in feces by guaiac fecal occult blood testing (gFOBT) has been proven to reduce CRC-related mortality.3 In recent years, fecal immunochemical testing (FIT) has become the preferred method of detecting fecal occult blood for CRC screening. FIT is more sensitive for the detection of CRC and its precursors.4,5 Besides, FIT is easier to handle than gFOBT.6 Consequently, screening participation rates increase.4,7,8 Also, FIT-analysis can be automated and quantitated. Quantitative FITs enable adjustment of cut-off to vary test-characteristics and match demand with available resources, in particular colonoscopy capacity.9

Men and women differ with respect to the prevalence of advanced colorectal neoplasia, with men having substantial higher prevalence of advanced adenomas and CRC than women.10,11 Repeated biennial gFOBT screening leads to a higher overall mortality reduction in men than in women.3 In fact, a recent gFOBT-based study showed that the prevalence of colorectal neoplasms was higher in men with a negative test than in women with a positive test.12 Moreover, male gender seems to be a stronger predictor of CRC than a positive gFOBT.13 Results from the national gFOBT screening program in Scotland showed a lower proportion of interval cancers for men compared to women.14 One study, using FIT, showed that men had higher positivity rates, as well as a higher detection rate.15 However, this study was limited as only FIT-positive (i.e. a fecal hemoglobin concentration >10 µg Hb/g feces) screenees underwent colonoscopy.

As more screening programs are being implemented worldwide, these gender differences become more apparent. Despite these differences, screening programs consistently use same strategies for both genders with regards to cut-off and screening intervals.10,16 Even though the use of different cut-offs in men and women would allow tailored screening strategies for each gender and improve CRC screening efficacy. Most studies on gender differences in CRC screening used gFOBT or FIT with one cut-off for both genders. Results were often based on assessing equal sensitivity of the test for men and women, thereby not taking into account gender differences regarding detection rate and miss rate of lesions in absolute numbers. Therefore, we aimed to illustrate the effect of gender tailored FIT screening including the detection and miss rates of advanced neoplasia.

Methods

The protocol of this population-based screening pilot (trialregister.nl; identifier NTR3549) has been described previously in detail.17,18 All authors had access to the study data and reviewed and approved the final manuscript.

Study population

Between June 2009 and July 2010, 6,600 asymptomatic individuals aged 50-75 years, living in the Amsterdam and Rotterdam regions were randomly selected from regional municipal administration registrations. They were invited for colonoscopy screening as primary screening modality or invited for computed tomography colonography. For the purpose of this manuscript only data of the population undergoing colonoscopy were used.

Individuals with a history of inflammatory bowel disease or CRC, as well as those who had undergone a full colonic examination in the past 5 years, those with an estimated life expectancy of <5 years, and subjects who were unable to give informed consent were excluded from the study. As there was no CRC screening program at the time of the trial in the Netherlands, the target population was screening-naive when first approached.

Fecal occult blood screening and colonoscopy

Eligible subjects who gave informed consent for colonoscopy screening were asked to complete one sample FIT (OC-sensor, Eiken, Japan) before colonoscopy. Participants were instructed to perform FIT at home, within 48 hours before the colonoscopy, but before starting the bowel preparation. No dietary restrictions were given. All patients underwent subsequent colonoscopy by experienced endoscopists. Research staff attended all colonoscopies and prospectively documented colonoscopy quality indicators and data on CRC and polyp detection.

Histology

Experienced pathologists classified all removed lesions as non-neoplastic, serrated polyp, adenoma (tubular, tubulovillous or villous) or carcinoma. Dysplasia was defined as low-grade or high-grade. Advanced adenomas were defined as an adenoma larger than 10 mm, an adenoma with villous histology (>25%) and/or an adenoma with high-grade dysplasia. Advanced neoplasia (AN) included both AA and CRC.

Statistical analysis

All screening participants who completed a FIT and subsequently underwent colonoscopy were included in the analysis. Baseline characteristics were described using descriptive statistics. The Chi-square test was used for comparing proportions of advanced neoplasia between men and women. The Mann-Whitney U test was used for non-parametric distributions. The sensitivity, specificity, positive and negative predictive value (PPV / NPV), and detection rate (DR) of advanced neoplasia were calculated for the most commonly used cut-offs; 10 (FIT10), 20 (FIT20), 30 (FIT30) and 40 (FIT40) µg Hb/g feces. These values correspond to 50, 100, 150 and 200 ng Hb/ml buffer. Following, sensitivity and specificity for fecal Hb-concentrations for all cutoffs between 0 and 100 µg Hb/g feces were calculated. Absolute numbers of detected and missed AN per 1000 subjects screened were calculated for men and women.

Results

Baseline characteristics and colonoscopy outcome

In total 1,256 invitees underwent FIT and colonoscopy, 638 men and 618 women. Men (61 years, IQR 56-66) were slightly older than women (60 years, IQR 55-65). Gender-specific findings at colonoscopy are described in Table 1. AN detection rate was slightly higher in men than in women, 10.6% (68/638) versus 8.3% (51/618) (p=0.146). CRC was detected in 5 (0.8%) men and in 3 (0.5%) women. No differences between men and women were seen in location of AN or number of AN per participant. In subjects with CRC, the median fecal hemoglobin concentration was 61 µg Hb/g feces (range 0-251 µg Hb/g) in men and 77 µg Hb/g feces (range 13-448 µg Hb/g) in women (p=0.76). In subjects with AN, men had a median fecal Hb concentration of 3.2 µg Hb/g feces (range 0-485 µg Hb/g) and women 2.6 µg Hb/g feces (IQR 0-670 µg Hb/g) (p=0.94).

Test characteristics

Performance characteristics of FIT for AN are provided in Table 2. Differences in test characteristics were not significant. At each of the pre-specified cut-offs, men had slightly higher positivity rates than women. The positivity rates ranged from 3.2% to 10.8% for the highest and lowest cut-off. The sensitivity for AN ranged from 40% (95% CI 29-52%) at FIT10 to 25% (95% CI 16-37%) at FIT40 in men, and from 35% (95% CI 24-49%) at FIT10 to 22% (95% CI 12-35%) at FIT40 in women. The specificity of FIT for AN tended to be lower in men when compared to women up to cut-offs of 20 µg Hb/g. The detection rate of AN was higher in men than women at all cut-offs. False positivity rates ranged from 1.1% (95% CI 0.5-2.3%) to 6.6% (95% CI 4.9-8.8%) for men and from 1.5% (95% CI

0.8-2.8%) to 5.5% (95% CI 4.0-7.6%) for women. True positivity rates ranged from 2.7% (95% CI 1.7-4.2%) to 4.2% (95% CI 2.9-6.1%) for men and from 1.8% (95% CI 1.0-3.2%) to 2.9% (95% CI 1.8-4.6%) for women. Sensitivity and specificity were calculated for the study population at multiple cut-offs ranging from 0 to 100 µg Hb/g feces (Figure 1). At an increasing cut-off, in both genders there is a relatively more rapid decline in sensitivity than an increase in specificity. Overall men had slightly higher sensitivities than women. For example, at a commonly used cut-off of 10 µg Hb/g feces women should have a lower cut-off to reach the same sensitivity and specificity as men.

Table 1. *Findings at colonoscopy in men and women.*

	Men	Women	p-value
	n = 638 (%)	n = 618 (%)	
Most advanced finding at colonoscopy			
no histology	17 (2.7)	3 (0.5)	
no abnormalities	303 (47.5)	357 (57.8)	
SSA < 10mm	19 (3.0)	17 (2.8)	
HP	82 (12.8)	77 (12.4)	
TA < 10mm	149 (23.4)	113 (18.3)	
TA ≥ 10 mm	21 (3.3)	16 (2.6)	
SSA ≥ 10mm	1 (0.2)	1 (0.2)	
TVA	32 (5.0)	25 (4.1)	
VA	2 (0.3)	0	
HGD	7 (1.1)	6 (0.9)	
CRC	5 (0.8)	3 (0.5)	
Total advanced neoplasia	68 (10.6)	51 (8.3)	0.15
Location of most advanced neoplasia**			
distal / proximal	51 (75) / 17 (25)	38 (75) / 13 (25)	0.95
Number of advanced neoplasia per participant**			
1 / >1	53 (77.9) / 15 (22.1)	42 (82.4) / 9 (17.6)	0.55

* no histology: removed polyp not retrieved for histology; SSA: sessile serrated adenoma; HP: hyperplastic polyp; TA: tubular adenoma; TVA: tubular villous adenoma; VA villous adenoma; HGD: high-grade dysplasia; CRC: colorectal cancer.

** only subjects with advanced neoplasia included (men n = 68 and women n = 51)

Detected lesions in absolute numbers

At all cut-offs, more lesions were detected as well as missed in men than in women (Figure 2 A, B). For all cut-offs the number needed to screen to identify one screenee with AN was higher in women than in men. It ranged from 38 to 56 subjects in women and from 24 to 38 subjects in men. Stepwise lowering cut-offs for men from FIT40 to respectively FIT30, FIT20 and FIT10 successively resulted in additional detection of 3, 0, and 7 AN. This required 8, 8, and 29 additional colonoscopies. Stepwise

lowering the cut-offs for women from FIT40 to FIT10 successively resulted in additional detection of 2, 4 and 1 AN. This required 2, 4 and 21 additional colonoscopies.

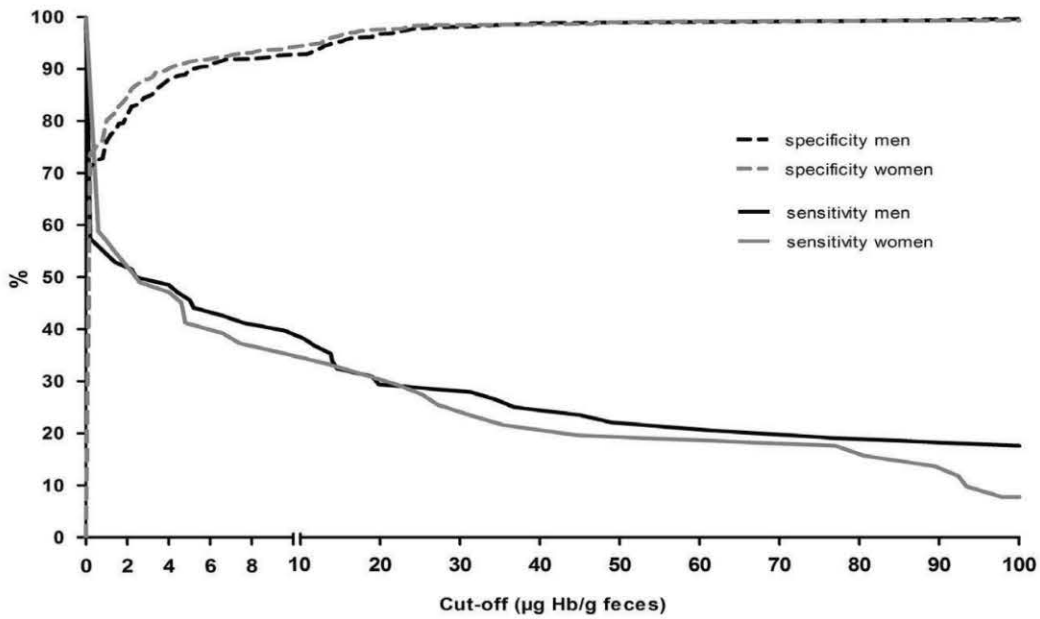


Figure 1. *Sensitivity and specificity for men and women for all cut-offs ranging from 0 to 100 µg Hb/g feces.*

Discussion

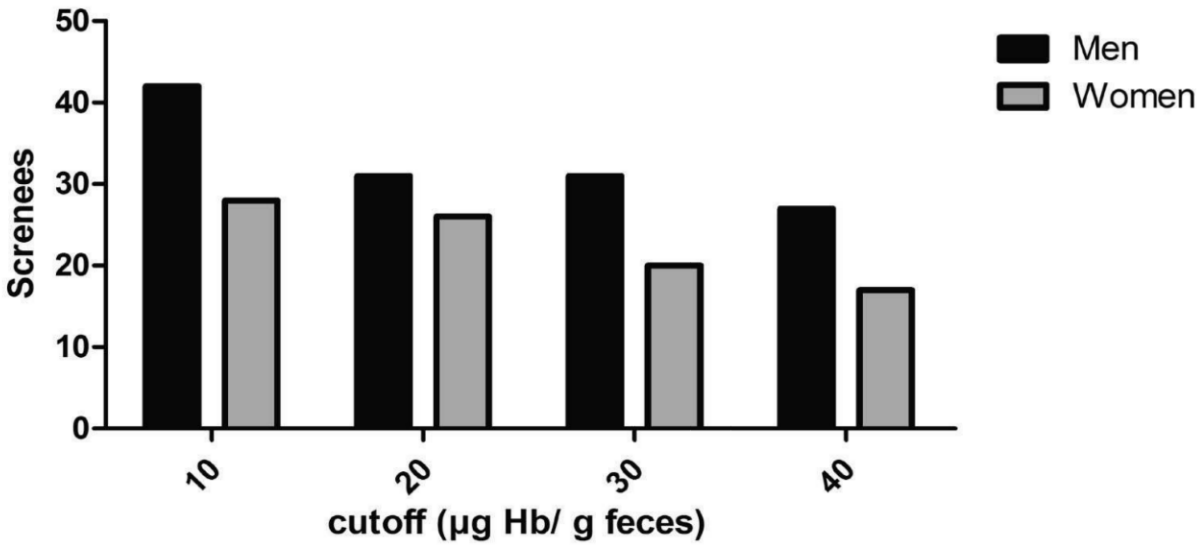
In this colonoscopy-based screening program, we evaluated gender differences with respect to the efficacy of FIT screening in average risk individuals. Furthermore, we illustrated the effect of using different cut-offs on a broad spectrum of screening outcomes. Our study demonstrated that FIT had a higher sensitivity and lower specificity for AN in men than in women. By increasing the cut-off a relative more rapid decline in sensitivity was found than an increase in specificity for both genders. Furthermore, FIT had an overall higher PPV in men. When looking at diagnostic yield in absolute numbers, men had higher detection rates and miss rates of AN than women at all cut-offs. This last finding is of particular interest as in current literature little attention has been given to gender specific miss rates of lesions.

A strength of this study is that this cohort was set in a population-based screening setting, making these results representative for average-risk screening populations. Also, as all participants underwent both colonoscopy and FIT, it is a very suitable population to demonstrate actual differences and to estimate the number of missed lesions. However, to appreciate our findings

Table 2. *Positivity rate, sensitivity, specificity, positive predictive value, negative predictive value, detection rate for men and women at different cut-offs.*

		Men		Women	
		%	(95%CI)	%	(95%CI)
FIT 10					
	PR	10.8	(8.6-13.5)	8.4	(6.5-10.9)
	sensitivity	39.7	(28.8-51.7)	35.3	(23.5-49.2)
	specificity	92.6	(90.2-94.5)	94.0	(91.7-95.7)
	PPV	39.1	(28.4-51.0)	34.6	(23.0-48.8)
	NPV	92.7	(90.4-94.7)	94.2	(91.9-95.8)
	DR	4.2	(2.9-6.1)	2.9	(1.8-4.6)
FIT 20					
	PR	6.3	(4.6-8.4)	5.0	(3.5-7.0)
	sensitivity	29.4	(19.8-41.2)	33.3	(21.8-47.2)
	specificity	96.4	(94.6-97.7)	97.5	(95.9-98.5)
	PPV	50.0	(35.0-65.0)	54.8	(37.4-71.1)
	NPV	92.0	(89.5-93.9)	94.2	(92.0-95.8)
	DR	3.1	(2.0-4.8)	2.8	(1.7-4.4)
FIT 30					
	PR	5.0	(3.6-7.0)	3.6	(2.4-5.3)
	sensitivity	29.4	(19.8-41.2)	25.5	(15.4-39.1)
	specificity	98.0	(96.3-98.8)	98.4	(97.0-99.2)
	PPV	62.5	(44.9-77.3)	59.0	(38.2-77.2)
	NPV	92.1	(89.6-94.0)	93.6	(91.4-95.3)
	DR	3.1	(2.9-4.8)	2.1	(1.2-3.6)
FIT 40					
	PR	3.8	(2.5-5.6)	3.2	(2.1-5.0)
	sensitivity	25.0	(16.1-36.6)	21.6	(12.4-34.9)
	specificity	98.8	(97.4-99.4)	98.4	(97.0-99.2)
	PPV	70.8	(50.2-85.4)	55.0	(33.6-74.7)
	NPV	91.7	(89.2-93.6)	93.3	(91.0-95.1)
	DR	2.7	(1.7-4.2)	1.8	(1.0-3.2)

A.



B.

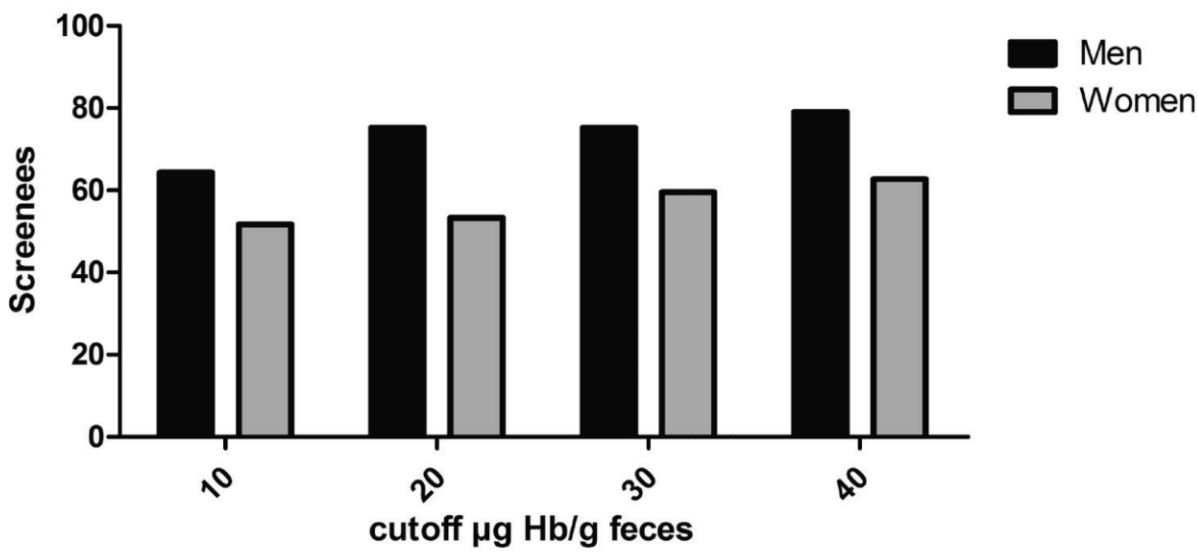


Figure 2. A. *Detected advanced neoplasia per 1000 screenees (for 100% participation in absolute numbers).* ***B.*** *Missed advanced neoplasia per 1000 screenees (for 100% participation in absolute numbers).*

some limitations need to be discussed. Firstly, this cohort consists of relatively small numbers and was not powered to detect differences in men and women. Another limitation is that only persons willing to undergo colonoscopy as primary screening method participated, which could have led to a selection bias resulting in a population that is not representative of FIT-participants. Only 22% of all invitees decided to participate in colonoscopy screening, while FIT-screening generally has much higher participation rates around 60% 19,20. Nevertheless, the population includes only screening-naive average risk subjects, and therefore we think that the risk of selection-bias is limited.
The introduction of fecal immunochemical testing was an important step forward in population-based colorectal cancer screening. FIT enables simple, low burden primary screening at relatively low costs and has a high uptake. For this reason, an increasing number of countries have implemented FIT-based screening programs or are in the process of doing so.21 This is mostly associated with a marked increase in colonoscopy demand. This asks for a strong focus on optimal use of limited resources.

Differences between men and women in terms of number of advanced lesions, location of lesions and fecal hemoglobin concentrations are becoming more evident. Dissimilarities in prevalence of AN between men and women have been well described with men having a substantial higher prevalence of AN than women.10,11 Consequently, research on tailored screening strategies become of significant importance. We are the first to describe the detection and miss rate of lesions in absolute numbers, showing that in men more lesions were both detected and missed for all cut-offs. This was especially the case at higher cut-offs. A previous Polish study showed that the number needed to screen in colonoscopy-based CRC screening to identify one screenee with advanced neoplasia was considerably higher in women than in men 10 Our data shows that these numbers also apply to colorectal cancer screening programs based on FIT. At each FIT cut-off, 14 to 18 more women needed to be screened to find one case of AN compared to men.

Differences in FIT screening between men and women can be explained by a combination of factors. It has been suggested that because men have a higher hemoglobin concentration in general, blood from bleeding polyps will contain more globin.22 As FIT specifically detects globin in feces, blood from these polyps could be detected more frequently in men. This is supported by the fact that differences in fecal Hb concentration have been found in men and women.15,23 A second explanation could be that women have more right-sided lesions, as it is known that fecal occult blood testing may not be as sensitive for proximal lesions as it is for distal lesions.14,22,24 Yet, our data did not show differences in location of AN between men and women. Another reason

for gender differences in FIT test-characteristics, could be the differences in colonic transit time between men and women, with women having slower transit times.25 A slower transit time could lead to more degradation of Hb and could decrease the likelihood of blood being detected by FIT.
An important question to be answered is how these results can be applied in colorectal cancer screening programs. Essentially, for gender-adjusted cut-offs in FIT-based CRC screening programs three scenarios are possible. These are the use of the same cut-off in both genders or, using a higher cut-off in men than women, or vice-versa. An increase of the cut-off for men compared to women can lead to a similar proportional sensitivity for detection of AN in both groups. As a consequence the difference in PPV between men and women would increase, with men having a substantial higher PPV. Also, a higher cut-off in men would lead to a further increase in miss rates of AN in men in absolute numbers and thus to a further increase in difference of miss rates in terms of absolute numbers of advanced lesions compared to women. Furthermore, using a lower cut-off for women would result in a higher rate of false-positive tests in women. The opposite strategy, i.e. increasing the cut-off for women compared to men, can lead to a similar miss rate in terms of absolute numbers, and to a similar PPV in both genders. It would however result in decreased sensitivity and detection rates for women. In this scenario a larger proportion of the colonoscopy capacity would be used for men. However, such a strategy could make sense given that men are at higher risk of AN and subsequently the development CRC.

Other gender-based CRC screening strategies besides adjusting the cut-off, include the use of different age ranges for screening, changing screening modality, or the use of different screening intervals. A German study showed that women reached equivalent levels of CRC-related mortality as men at a 4 to 8 years higher age 26. Gender differences in other screening modalities, such as colonoscopy, sigmoidoscopy, fecal biomarkers and fecal DNA, have not yet been extensively investigated. However, using different methods or combinations of tests for men and women could optimize screening efficacy and should be further investigated. With regard to gender differences in patient-education there is still much to gain. Information on miss rates of advanced lesions is an important issue in client information. At present, men and women are informed in the same manner about FIT-based CRC screening. These results helps to accurately inform the client about the gender-dependent risk of miss rates and detection of advanced lesions in a FIT based CRC screening program.

To conclude, colorectal cancer screening using FIT with the same cut-off for both genders results in a higher sensitivity and lower specificity for advanced neoplasia in men than in women. In absolute

numbers more advanced neoplasia are detected and missed in men for all cut-offs. Following, tailored cut-off based on gender could either level sensitivity in men and women by using a lower cut-off in women, or level the amount of missed lesions when using a lower cut-off in men. Adjusting cut-offs based on gender can contribute to the efficacy of FIT-based CRC screening programs and optimize the use of available endoscopy resources. In addition, individuals invited to attend a FIT based CRC screening should be informed accordingly about these gender differences.

References

1. Center MM, Jemal A, Ward E. International trends in colorectal cancer incidence rates. Cancer Epidemiol Biomarkers Prev 2009; 18(6): 1688-94.
2. Edwards BK, Ward E, Kohler BA, et al. Annual report to the nation on the status of cancer, 1975-2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. Cancer 2010; 116(3): 544-73.
3. Shaukat A, Mongin SJ, Geisser MS, et al. Long-term mortality after screening for colorectal cancer. N Engl J Med 2013; 369(12): 1106-14.
4. European Colorectal Cancer Screening Guidelines Working G, von Karsa L, Patnick J, 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.
5. Hol L, Wilschut JA, van Ballegooijen M, et al. Screening for colorectal cancer: random comparison of guaiac and immunochemical faecal occult blood testing at different cut-off levels. Br J Cancer 2009; 100(7): 1103-10.
6. 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.
7. Brenner H, Tao S. Superior diagnostic performance of faecal immunochemical tests for haemoglobin in a head-to-head comparison with guaiac based faecal occult blood test among 2235 participants of screening colonoscopy. Eur J Cancer 2013.
8. 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.
9. Kuipers EJ, Rosch T, Bretthauer M. Colorectal cancer screening--optimizing current strategies and new directions. Nat Rev Clin Oncol 2013; 10(3): 130-42.
10. Regula J, Rupinski M, Kraszewska E, et al. Colonoscopy in colorectal-cancer screening for detection of advanced neoplasia. N Engl J Med 2006; 355(18): 1863-72.
11. Brenner H, Hoffmeister M, Stegmaier C, Brenner G, Altenhofen L, Haug U. Risk of progression of advanced adenomas to colorectal cancer by age and sex: estimates based on 840,149 screening colonoscopies. Gut 2007; 56(11): 1585-9.
12. Brenner H, Hoffmeister M, Birkner B, Stock C. Men with negative results of guaiac-based fecal occult blood test have higher prevalences of colorectal neoplasms than women with positive results. Int J Cancer 2014; 134(12): 2927-34.
13. Ferlitsch M, Heinze G, Salzl P, et al. Sex is a stronger predictor of colorectal adenoma and advanced adenoma than fecal occult blood test. Med Oncol 2014; 31(9): 151.
14. Steele RJ, McClements P, Watling C, et al. Interval cancers in a FOBT-based colorectal cancer population screening programme: implications for stage, gender and tumour site. Gut 2012; 61(4): 576-81.
15. Kapidzic A, van der Meulen MP, Hol L, et al. Gender differences in fecal immunochemical test performance for early detection of colorectal neoplasia. Clin Gastroenterol Hepatol 2015.
16. Ferlitsch M, Reinhart K, Pramhas S, et al. Sex-specific prevalence of adenomas, advanced adenomas, and colorectal cancer in individuals undergoing screening colonoscopy. JAMA 2011; 306(12): 1352-8.
17. de Wijkerslooth TR, de Haan MC, Stoop EM, et al. Study protocol: population screening for colorectal cancer by colonoscopy or CT colonography: a randomized controlled trial. BMC Gastroenterol 2010; 10: 47.

18. de Wijkerslooth TR, Stoop EM, Bossuyt PM, et al. Immunochemical fecal occult blood testing is equally sensitive for proximal and distal advanced neoplasia. Am J Gastroenterol 2012; 107(10): 1570-8.
19. de Wijkerslooth TR, de Haan MC, Stoop EM, et al. Burden of colonoscopy compared to non-cathartic CT-colonography in a colorectal cancer screening programme: randomised controlled trial. Gut 2012; 61(11): 1552-9.
20. Kapidzic A, Grobbee EJ, Hol L, et al. Attendance and yield over three rounds of population-based fecal immunochemical test screening. Am J Gastroenterol 2014; 109(8): 1257-64.
21. Schreuders EH, Ruco A, Rabeneck L, et al. Colorectal cancer screening: a global overview of existing programmes. Gut 2015.
22. Brenner H, Haug U, Hundt S. Sex differences in performance of fecal occult blood testing. Am J Gastroenterol 2010; 105(11): 2457-64.
23. Auge JM, Pellise M, Escudero JM, 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.
24. Chacko L, Macaron C, Burke CA. Colorectal Cancer Screening and Prevention in Women. Dig Dis Sci 2015.
25. Meier R, Beglinger C, Dederding JP, et al. Influence of age, gender, hormonal status and smoking habits on colonic transit time. Neurogastroenterol Motil 1995; 7(4): 235-8.
26. Brenner H, Hoffmeister M, Arndt V, Haug U. Gender differences in colorectal cancer: implications for age at initiation of screening. Br J Cancer 2007; 96(5): 828-31.


chapter
10

Second look colonoscopies and the impact on capacity in FIT-based colorectal cancer screening

Esmée J. Grobbee, Atija Kapidzic, Anneke J. van Vuuren, Monique van Leerdam, Iris Lansdorp-Vogelaar, Caspar W.N. Looman, Marco J. Bruno, Ernst J. Kuipers, Manon C.W. Spaander

American Journal of Gastroenterology, 2015

Abstract

Objectives Fecal immunochemical testing (FIT) and colonoscopy are tandem procedures in colorectal cancer screening (CRC). A positive FIT predicts advanced neoplasia (AN), which requires endoscopic detection and removal. En bloc or piecemeal resection of AN is associated with a significant rate of residual or recurrent neoplasia. Second look colonoscopies are indicated to assess completeness of removal of AN. These colonoscopies can make a substantial demand on colonoscopy capacity and health care system. This study is the first to evaluate the demand and risk factors for second look colonoscopy in FIT CRC screening.

Methods All colonoscopies after a positive FIT, in subjects aged 50-74 years approached for 3 rounds of FIT screening, were prospectively registered. Second look colonoscopies were defined as any colonoscopy within one year following a colonoscopy after positive FIT.

Results Out of 1215 FIT-positive screenees undergoing colonoscopy, 105 (8.6%) patients underwent a second look colonoscopy of whom 30 (2.5%) underwent more than one (range 2-9) leading to a total of 149 (12.3%) additional colonoscopies. Main reasons for second look colonoscopies were assessment of complete AN removal (41.9%), and need for additional polypectomy (34.3%). Risk factors were advanced adenomas and poor bowel preparation ($p<0.001$). High fecal hemoglobin concentration was the only predictor of a second look colonoscopy before the index-colonoscopy ($p<0.001$).

Conclusions Second look colonoscopies have substantial impact on colonoscopy resources, increasing the demand with 12%. The main reasons for these second-look colonoscopies were previous incomplete polypectomy and control of completeness of removal of neoplastic lesions. A high fecal hemoglobin concentration as measured by FIT can help to identify patients at risk of a second look colonoscopy.

Introduction

Colorectal cancer (CRC) is a major cause of mortality and morbidity worldwide 1. Endoscopic detection and removal of adenomatous polyps reduces CRC-related mortality[2,3]. Colonoscopy is widely appreciated as optimal test for detection and removal of adenomas[4,5]. At present more and more CRC-screening programs are implemented worldwide. Recent EU guidelines recommend fecal immunochemical occult blood testing (FIT) as a tool for primary screening followed by colonoscopy in case of positive FIT[6,7]. However colonoscopic examination has a low but not negligible miss rate for cancers and adenomatous polyps[8]. In one recent study, persistent adenomatous tissue was observed in 10.1% of cases after colonoscopic polypectomy[9]. Missed and incompletely resected lesions are recognized as important contributors to interval colorectal cancers[10,11]. A so-called second look colonoscopy is recommended in case of doubt about missed neoplastic lesions, completeness of removal of lesions, or after an incomplete colon examination[12,13]. The positive predictive value of FIT for advanced neoplasia varies around 35-45%[14-16]. En bloc, and especially piecemeal resection of advanced adenomas is associated with a relatively high rate of local residual or recurrent neoplasia[12,13,17,18]. One may therefore hypothesize that FIT screening may be associated with a significant rate of second look colonoscopies[14-16]. Although these colonoscopies can have a substantial impact on the required colonoscopy capacity and health care system, little is known about the demand for second look colonoscopy in a FIT screening program. This is to our knowledge the first study to assess the number and indications of second look colonoscopy in in a FIT-based CRC screening program and to identify patients at risk for a second look colonoscopy.

Methods

Patients

Details about the design of this on-going population-based CRC screening program have been described previously[14,15]. In short, demographic data of all individuals between 50-74 years living in the southwest of the Netherlands were obtained from municipal population registers. In this screening program the OC-sensor FIT (Eiken, Japan) was used over multiple rounds with a maximum of three rounds. Intervals between rounds varied from one to three years. Individuals with a history of inflammatory bowel disease or CRC, symptomatic patients, as well as those who had undergone a colonoscopy, sigmoidoscopy or barium contrast enema in the last 3 years, those with an estimated life expectancy of less than 5 years, and subjects who were unable to give informed consent, were excluded from the study. All patients with a positive FIT, defined by a hemoglobin (Hb) concentration

of ≥ 50 ng/ml which corresponds to ≥ 10 µg Hb/g feces, were referred for colonoscopy. Colonoscopies were performed in 18 non-academic centers and in 1 academic center.

Data collection

Our primary endpoint was to assess the number and indications of second look colonoscopies after the first screening colonoscopy following a positive FIT. Second-look colonoscopy was defined as any secondary endoscopic procedure of the colon indicated within one year after the first screening colonoscopy, regardless of the endpoint reached, as often a second look colonoscopy was limited to the area where previous neoplastic lesions were removed[19]. In addition we looked for predictive factors to identify patients at risk. Predictive factors included age, sex, socio-economic status (low, average, high), bowel preparation, use of sedation, use of buscopan, type of endoscopist, type of hospital, fecal hemoglobin concentration, and presence of advanced neoplasia.

Colonoscopy data were prospectively registered using a standardized endoscopy report completed after the procedure by the endoscopist. The following variables were systematically assessed: sedation (midazolam, fentanyl, propofol, none), level of bowel preparation (poor: < 90% of mucosa visible, medium: 90-99% of mucosa visible, good: 100% of mucosa visible), cecal intubation, detection of polyps or other lesions, and removal of polyps. Endoscopists were categorized as gastroenterologists, gastroenterology fellows, internists or nurse-endoscopists. Advanced neoplasia was defined as an adenoma of 10mm or larger, an adenoma with 25% or more villous histology or with high-grade dysplasia, and CRC.

The overall quality of colonoscopy was evaluated based on indicators as defined by Rex et al 20. The cecal intubation rate (CIR) was defined as the proportion of colonoscopies in which the cecum was visualized. The adenoma detection rate (ADR) was defined as the number of colonoscopies that revealed at least one adenoma divided by the total number of colonoscopies.

Statistical analysis

Descriptive data were reported as proportions or means with the standard deviation. For non-normally distributed data the median and interquartile range (IQR) were given. Chi-Square tests were used to analyze categorical data; continuous data were analyzed using Student's t-tests and Mann-Whitney U in case of a non-parametric distribution. Univariate logistic regression models were used to calculate odds ratios (OR) and 95% confidence intervals (CI) to assess the risk of a second look colonoscopy. In case of a p-value < 0.20 variables were included in multivariate stepwise

backward regression analysis and described as the full model. All variables that remained significant using the stepwise backward method were shown in a final model. In multivariate analysis a full case analysis was performed. A two-sided p-value of < 0.05 was considered statistically significant. Statistical analysis was performed using IBM SPSS version 21.0.

Results

Patient characteristics

A total of 1215 patients, 698 men and 517 women, with a median age of 63 years (IQR 57-68 years), underwent a colonoscopy following a positive FIT. Patients' characteristics are summarized in Table 1. The median fecal hemoglobin concentration was 29 µg Hb/g feces (IQR 15 – 86 µg Hb/g).

Table 1. *Baseline demographics and colonoscopy characteristics (n=1215)*

Age, mean (years, IQR)	63	(57-68)
Male gender (n,%)	698	(57.4)
Sedation (n, %)		
Midazolam	348	(28.6)
Fentanyl	22	(1.8)
Midazolam and fentanyl	632	(52.0)
Propofol	2	(0.2)
No sedation	172	(14.2)
Not reported	39	(3.2)
Use of buscopan (n, %)	484	(39.8)
Endoscopist (n, %)		
Gastroenterologist	911	(75.0)
Internist	66	(5.4)
Gastroenterology fellow	27	(2.2)
Nurse-endoscopist	156	(12.8)
Not reported	55	(4.5)
Bowel preparation (n, %)		
Good	885	(72.8)
Medium	241	(19.8)
Poor	29	(2.4)
Not reported	60	(4.9)
Hospital (n, %)		
Academical	255	(21.0)
Non-academical	960	(79.0)
Faecal Hb concentration (µg Hb/g feces, IQR)	29	(15-86)

Colonoscopy characteristics

Of the 1215 colonoscopies, more than half (52.0%) were performed under conscious sedation using both midazolam and fentanyl. Buscopan was used in 484 (39.8%) of the cases. Colonoscopies were performed by gastroenterologists (75.0%), nurse-endoscopists (12.8%), internists (5.4%), and fellows (2.2%). The bowel was adequately cleansed in 1126 (97%) of 1155 patients. The cecum was reached in 97.3% of the index colonoscopies. The overall adenoma detection rate was 55%. Adverse events within 30 days occurred in 36 colonoscopies (3%), consisting mainly of mild bleedings (1.8%) managed during the index colonoscopy. Other adverse events were a decrease in saturation (0.4%) and blood pressure (0.2%) during colonoscopy. One iatrogenic perforation (0.09%) occurred, this was noted after colonoscopy in the sigmoid and required surgical removal of the perforated tissue.

Table 2. *Indications for second look colonoscopy (n=105).*

Indication	**n**	**(%)***
Control of completeness of removal of neoplastic lesion	46	(43.8)
piecemeal resection	*19*	*(18.1)*
resected polyps with positive histological margins	*14*	*(13.3)*
high grade dysplasia or CRC in pathology	*9*	*(8.5)*
doubt about endoscopic resection	*4*	*(3.8)*
Additional polypectomy	34	(32.4)
polyp size too large	*20*	*(19.0)*
*suspected high grade dysplasia or CRC***	*5*	*(6.7)*
lack of time	*6*	*(5.7)*
complex location	*2*	*(1.9)*
procedure too painful	*1*	*(0.8)*
Poor bowel preparation	14	(13.3)
Marking adenoma / CRC	4	(3.8)
Anticoagulant drugs	2	(1.9)
Incomplete colonoscopy ***	4	(3.8)
Obstructing CRC	1	(1.0)

* does not add up to 100% due to rounding

** biopsies were taken at initial colonoscopy

*** includes: looping, diverticulosis, diverticulitis

Second look colonoscopies

A total of 105 (8.6%) patients underwent a second look colonoscopy within one year, with a median time between the index colonoscopy of 63 days (IQR 35-101 days). The most frequently reported indications for second look colonoscopy were assessment of completeness of removal of a neoplastic lesion (43.8%) and additional polypectomy for various reasons (32.4%) (Table 2). Remaining indications were poor bowel preparation (13.3%), pre-surgical submucosal marking of

an adenoma or malignancy (3.8%), and anticoagulant use (1.9%). One patient with a obstructing CRC underwent a second look colonoscopy to examine the complete colon after surgical removal of the CRC.

Characteristics of the most advanced polyps removed in patients that underwent second-look colonoscopy are described in Table 3. The majority of these polyps were located in the recto-sigmoid (61%) and were relatively large (median 15 mm, IQR 10-25 mm).

Table 3. *Characteristics of most advanced polyps in patients that underwent second look colonoscopy (n = 80*)*

Size (in mm, median, IQR)	15	(10-25)
Location (n, %)		
cecum	*8*	*(10.0)*
ascendens	*8*	*(10.0)*
hepatic flexure	*2*	*(2.5)*
transversum	*6*	*(7.5)*
splenic flexure	*0*	*(0)*
descendens	*4*	*(5.0)*
sigmoid	*41*	*(50.0)*
rectum	*9*	*(11.2)*
*multiple locations***	*2*	*(3.8)*
Histology (n, %)		
tubulair adenoma	*21*	*(26.2)*
(tubulo)villous adenoma	*26*	*(32.5)*
high grade dysplasia	*23*	*(28.8)*
adenocarcinoma	*10*	*(12.5)*

* includes all second look colonoscopies after additional polypectomy or for control of completeness of removal of lesions

** 2 patients had multiple polyps as a reason for second look colonoscopy (26 and 38 polyps respectively); most advanced lesion is used for histology

Thirty patients (28.8%) underwent more than one follow-up colonoscopy (range 2-9 colonoscopies) leading to a total of 149 (12.3%) additional colonoscopies after the screening colonoscopy. The main indications for these subsequent colonoscopies were control for completeness of removal of a neoplasic lesion and additional polypectomy. In 61% of the colonoscopies with failed cecal intubation the decision was made not to repeat the colonoscopy and to refer the patient for CT-colonography.

Factors associated with second look colonoscopy

Predictors for a second look colonoscopy in the univariate analysis are shown in Table 4. Bowel preparation and advanced neoplasia were the only significant predictors for a second look

colonoscopy after multivariate analysis. The only predictive factor before the initial colonscopy was FIT Hb-concentration (Figure 1).

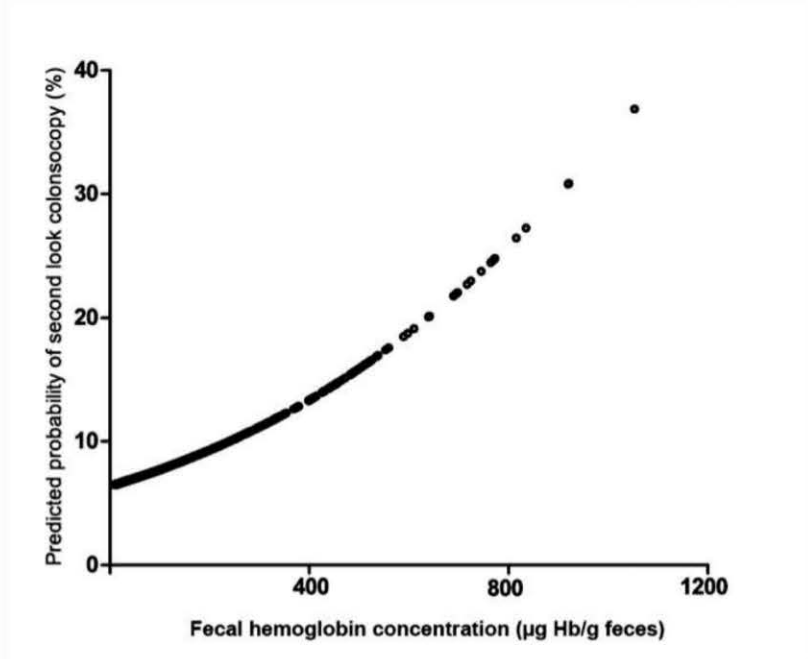


Figure 1. *Predicted probability of the risk of a second-look colonoscopy for fecal hemoglobin (Hb) level.*

Discussion

In this population FIT- based screening program we assessed the number and indications of second-look colonoscopies. The results of our study indicate that 8.6% of our study population required a second look colonoscopy within one year after the initial screening colonoscopy. In 2.5% of the patients more than one colonoscopy was performed after the initial procedure. Together, this led to a 12.3% increase in colonoscopy demand compared to the volume of primary colonoscopies after positive FIT. In 76% of the patients a second look colonoscopy was performed for assessment of completeness of removal of neoplastic lesions or for further polypectomy of large or multiple lesions. Significant predictors for a second look colonoscopy were presence of advanced neoplasia and poor bowel preparation. The only significant predictor prior to the index colonoscopy was a high fecal hemoglobin concentration.

While second look colonoscopies can have a substantial impact on colonoscopy demand, current knowledge on the volume of second look colonoscopies is limited. As more screening programs are implemented worldwide, it becomes important to estimate the number of colonoscopies needed for CRC screening. To our knowledge this is the first study to evaluate the need for second look

colonoscopies, as previous literature has mainly focused on failed cecal intubation as a cause for a second colonoscopy[19,21,22]. In our cohort cecal intubation failed in 3% of the colonoscopies. Next, many studies have suggested that a poor bowel preparation is a frequent reason for colonoscopy failure[23,24]. We found that, although a poor bowel preparation almost always leads to a repeat procedure, it is an infrequent cause of second look colonoscopy.

Previous FIT-based screening cohort studies have shown that FIT-positive screenees have a 35-45% incidence of advanced adenomas[15-17]. The removal of these adenomas is complex, time consuming and often followed by a second look colonoscopy to assess residual neoplastic tissue[25,26]. We found that in over two-thirds of our patients a second look colonoscopy was performed because of an incomplete polypectomy or to examine the polypectomy scar for residual neoplastic tissue. According to literature data, a second look colonoscopy for the control after polypectomy occurs in around 1% or 2% of the procedures with endoscopic treatment and is usually limited to the previously treated area[19]. Our results indicate a much higher rate of second look colonoscopies. These findings are supported by the fact that significantly more advanced neoplasia were found in the patients undergoing a second look colonoscopy. A second look colonoscopy is recommended to be performed within 2 to 6 months or at least within 1 year after piecemeal resection[17,18]. Our findings are in line with these guidelines as half of the second look colonoscopies were performed within 3 months following the index colonoscopy. It should be noted that in only 20% of the patients residual tissue was found at second look colonoscopy.

Advanced colorectal neoplasia is associated with a higher FIT hemoglobin concentration, and accordingly a higher Hb concentration leads to a higher positive predictive value for AN[27-30]. These results are consistent with our findings and this readily explains the relation between a high fecal hemoglobin concentration and the need for a second look colonoscopy. Our findings can be of clinical importance and guidance for endoscopists and patients. Currently the physician is only informed about the qualitative result of the test, i.e. a negative or positive result and not about the quantative result. However, the exact value of the fecal hemoglobin concentration could be used for estimating the chance of finding advanced pathology at colonoscopy. Since a high fecal hemoglobin concentration is indicative of more advanced pathology, it may help the endoscopists to be prepared for a more difficult procedure and inform the patient accordingly. For example, patients with very high concentrations can be scheduled at the program of endoscopists with experience for removal of advanced lesions.

Our study has several limitations. Although the data were prospectively collected, the number of second look colonoscopies were retrospectively analyzed which could lead to possible underreporting of the actual number of colonoscopies. Secondly, we lacked information regarding the colonoscopy experience of each endoscopist. However, both CIR and ADR were above standards as required for CRC screening[31,32]. Thirdly, many colonscopies were performed without buscopan, as the use of buscopan is not common practice in the Netherlands and its use is mainly based on the preference of the endoscopists. Finally, in the fecal samples with very high concentrations of Hb, a prozone effect could have occurred. This could lead to measured values that are lower than the actual concentration in the sample in case of very high concentrations[33]. Such a prozone effect could lead to an underestimation of the true height of the fecal Hb level for values above 200 µg Hb/g feces.

In conclusion, we found that 8.6% of patients that underwent colonoscopy after a positive FIT undergo a second look colonoscopy, with the total number of procedures ranging from 2 to 9 colonoscopies per patient. This leads to an extra 12% demand for screening colonoscopy. In over two-thirds of the patients a second look colonoscopy was performed for control of completeness of removal of neoplastic lesions or for additional polypectomy. FIT Hb-concentration was the only significant predictor prior to the screening colonoscopy and could be of clinical importance and guidance for endoscopists and patients. Our results show that second look colonoscopies have a substantial influence on colonoscopy burden in a FIT-based screening setting and should be taken into account when estimating colonoscopy capacity.

References

1. Edwards BK, Ward E, Kohler BA, et al. Annual report to the nation on the status of cancer, 1975-2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. Cancer 2010; 116(3): 544-73.
2. Winawer SJ, Zauber AG, Ho MN, et al. Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. N Engl J Med 1993; 329(27): 1977-81.
3. Zauber AG, Winawer SJ, O'Brien MJ, et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. N Engl J Med 2012; 366(8): 687-96.
4. Kuipers EJ, Rosch T, Bretthauer M. Colorectal cancer screening--optimizing current strategies and new directions. Nat Rev Clin Oncol 2013; 10(3): 130-42.
5. Chen SC, Rex DK. Endoscopist can be more powerful than age and male gender in predicting adenoma detection at colonoscopy. Am J Gastroenterol 2007; 102(4): 856-61.
6. European Colorectal Cancer Screening Guidelines Working G, von Karsa L, Patnick J, 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.
7. 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.
8. van Rijn JC, Reitsma JB, Stoker J, Bossuyt PM, van Deventer SJ, Dekker E. Polyp miss rate determined by tandem colonoscopy: a systematic review. Am J Gastroenterol 2006; 101(2): 343-50.
9. Pohl H, Robertson DJ. Colorectal cancers detected after colonoscopy frequently result from missed lesions. Clin Gastroenterol Hepatol 2010; 8(10): 858-64.
10. Pohl H, Srivastava A, Bensen SP, et al. Incomplete polyp resection during colonoscopy-results of the complete adenoma resection (CARE) study. Gastroenterology 2013; 144(1): 74-80 e1.
11. le Clercq CM, Sanduleanu S. Interval colorectal cancers: what and why. Curr Gastroenterol Rep 2014; 16(3): 375.
12. Buchner AM, Guarner-Argente C, Ginsberg GG. Outcomes of EMR of defiant colorectal lesions directed to an endoscopy referral center. Gastrointest Endosc 2012; 76(2): 255-63.
13. Khashab M, Eid E, Rusche M, Rex DK. Incidence and predictors of "late" recurrences after endoscopic piecemeal resection of large sessile adenomas. Gastrointest Endosc 2009; 70(2): 344-9.
14. Hol L, van Leerdam ME, van Ballegooijen M, et al. Screening for colorectal cancer: randomised trial comparing guaiac-based and immunochemical faecal occult blood testing and flexible sigmoidoscopy. Gut 2010; 59(1): 62-8.
15. van Roon AH, Goede SL, van Ballegooijen M, et al. Random comparison of repeated faecal immunochemical testing at different intervals for population-based colorectal cancer screening. Gut 2013; 62(3): 409-15.
16. Denters MJ, Deutekom M, Bossuyt PM, Stroobants AK, Fockens P, Dekker E. Lower risk of advanced neoplasia among patients with a previous negative result from a fecal test for colorectal cancer. Gastroenterology 2012; 142(3): 497-504.
17. Winawer SJ, Zauber AG, Fletcher RH, et al. Guidelines for colonoscopy surveillance after polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer and the American Cancer Society. Gastroenterology 2006; 130(6): 1872-85.
18. Lieberman DA, Rex DK, Winawer SJ, et al. Guidelines for colonoscopy surveillance after screening and polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer. Gastroenterology 2012; 143(3): 844-57.

19. Rey JF, Lambert R. Second look colonoscopy: indication and requirements. Dig Endosc 2009; 21 Suppl 1: S47-9.
20. Rex DK, Petrini JL, Baron TH, et al. Quality indicators for colonoscopy. Gastrointest Endosc 2006; 63(4 Suppl): S16-28.
21. Shah HA, Paszat LF, Saskin R, Stukel TA, Rabeneck L. Factors associated with incomplete colonoscopy: a population-based study. Gastroenterology 2007; 132(7): 2297-303.
22. Hsu CM, Lin WP, Su MY, Chiu CT, Ho YP, Chen PC. Factors that influence cecal intubation rate during colonoscopy in deeply sedated patients. J Gastroenterol Hepatol 2012; 27(1): 76-80.
23. Hautefeuille G, Lapuelle J, Chaussade S, et al. Factors related to bowel cleansing failure before colonoscopy: Results of the PACOME study. United European Gastroenterol J 2014; 2(1): 22-9.
24. Froehlich F, Wietlisbach V, Gonvers JJ, Burnand B, Vader JP. Impact of colonic cleansing on quality and diagnostic yield of colonoscopy: the European Panel of Appropriateness of Gastrointestinal Endoscopy European multicenter study. Gastrointest Endosc 2005; 61(3): 378-84.
25. Hori K, Uraoka T, Harada K, et al. Predictive factors for technically difficult endoscopic submucosal dissection in the colorectum. Endoscopy 2014.
26. Tholoor S, Tsagkournis O, Basford P, Bhandari P. Managing difficult polyps: techniques and pitfalls. Ann Gastroenterol 2013; 26(2): 114-21.
27. Hol L, Wilschut JA, van Ballegooijen M, et al. Screening for colorectal cancer: random comparison of guaiac and immunochemical faecal occult blood testing at different cut-off levels. Br J Cancer 2009; 100(7): 1103-10.
28. Crotta S, Segnan N, Paganin S, Dagnes B, Rosset R, Senore C. High rate of advanced adenoma detection in 4 rounds of colorectal cancer screening with the fecal immunochemical test. Clin Gastroenterol Hepatol 2012; 10(6): 633-8.
29. Ciatto S, Martinelli F, Castiglione G, et al. Association of FOBT-assessed faecal Hb content with colonic lesions detected in the Florence screening programme. Br J Cancer 2007; 96(2): 218-21.
30. Digby J, Fraser CG, Carey FA, et al. Faecal haemoglobin concentration is related to severity of colorectal neoplasia. J Clin Pathol 2013; 66(5): 415-9.
31. Rex DK, Petrini JL, Baron TH, et al. Quality indicators for colonoscopy. Am J Gastroenterol 2006; 101(4): 873-85.
32. Rex DK, Bond JH, Winawer S, et al. Quality in the technical performance of colonoscopy and the continuous quality improvement process for colonoscopy: recommendations of the U.S. Multi-Society Task Force on Colorectal Cancer. Am J Gastroenterol 2002; 97(6): 1296-308.
33. –Vaananen P, Tenhunen R. Rapid immunochemical detection of fecal occult blood by use of a latex-agglutination test. Clin Chem 1988; 34(9): 1763-6.


chapter
11

Immunochemical fecal occult blood testing to screen for colorectal cancer: can the screening interval be extended?

Ulrike Haug, Esmée J. Grobbee, Iris Lansdorp-Vogelaar, Manon C.W. Spaander, Ernst Kuipers

Gut, 2015

Abstract

Objective Colorectal cancer (CRC) screening programs based on fecal immunochemical testing for hemoglobin (FIT) typically use a screening interval of 2 years. We aimed to estimate how alternative FIT strategies that use a lower than usual positivity threshold followed by a longer screening interval compare to conventional strategies.

Methods We analysed longitudinal data of 4523 Dutch individuals (50–74 years at baseline) participating in round I of a one-sample FIT screening program, of which 3427 individuals also participated in round II after 1-3 years. The cohort was followed until 2 years after round II. In both rounds, a cut-off level of ≥50ng hemoglobin (Hb)/ml (corresponding to 10µg Hb/g feces) was used, representing the standard scenario. We determined the cumulative positivity rate (PR) and the numbers of subjects diagnosed with advanced adenomas (N_AdvAd) and early stage CRC (N_earlyCRC) in the cohort over two rounds of screening (standard scenario) and compared it to hypothetical single round screening with use of a lower cut-off and omission of the second round (alternative scenario).

Results In the standard scenario, the cumulative (i.e. round I and II combined) PR, N_AdvAd, and N_earlyCRC were 13%, 180, and 26, respectively. In alternative scenarios using a cutoff level of respectively ≥11ng/ml and ≥22ng/ml (corresponding to 2 and 4µg Hb/g feces), the PRs were 18% and 13%, the N_AdvAd were 180 and 162, and the N_earlyCRC ranged between 22-27 and 22-26.

Conclusion The diagnostic yield of FIT screening using a lowered positivity threshold in combination with an extended screening interval (up to 5 years) may be similar to conventional FIT strategies. This justifies and motivates further research steps in this direction.

Introduction

Colorectal cancer (CRC) is the third most common cancer and cause of cancer-related deaths worldwide, with more than 1.2 million new cases and more than 600,000 deaths per year[1]. Randomized controlled trials have demonstrated that biennial screening with guaiac-based fecal occult blood testing (gFOBT) reduces CRC mortality by 15%[2].

Since the conduct of these trials, a substantial body of evidence has shown that the diagnostic performance of fecal immunochemical testing for hemoglobin (FIT) is superior to gFOBT screening[3-6]. Furthermore, FIT screening is associated with higher participation rates than gFOBT and has analytical advantages such as no cross-reactivity with dietary constituents and the option of measuring fecal hemoglobin levels quantitatively[5,6]. Accordingly, the European guidelines for quality assurance in CRC screening recommend screening with FIT in preference to gFOBT[7]. Current CRC screening programs based on FIT are typically designed analogously to gFOBT strategies[8]. That means that they use a screening interval of two years and a positivity threshold yielding a specificity well above 90%.

However, studies reporting on repeated rounds of FIT screening consistently showed that the diagnostic yield of the first re-screening round is lower as compared to the initial screening round[9-11]. We further showed that the diagnostic yield in a second round of FIT screening is the same when using a 1-, 2-, or 3-year interval[9]. When combining these observations with the fact that FIT – given its quantitative nature - offers flexibility to select the positivity threshold, it leads to the idea of considering alternative strategies for FIT screening. The positivity threshold at the initial round could be lowered, thereby increasing sensitivity at the cost of decreasing specificity. Subsequently, the interval to the second round could be significantly increased.

Although strategies with a lower positivity threshold in combination with a longer interval may have advantages, for instance regarding the organizational effort, it needs to be considered carefully whether there are disadvantages, e.g. regarding the diagnostic yield and the colonoscopy load. We aimed to explore and estimate how such alternative FIT strategies compare to conventional FIT screening in terms of positivity rate and thus colonoscopy demand, detection of advanced adenomas and CRC detection.

Methods

Study overview

To address the research question, we analysed longitudinal data of an ongoing population-based CRC screening study that started in 2006 and aimed to investigate the diagnostic yield and adherence patterns over multiple rounds of FIT screening. As illustrated in Figure 1, we focused on the first and the second round of this study and considered the positivity rate and the diagnostic yield of these two rounds cumulatively. We then compared this standard scenario to various alternative scenarios that assumed use of a lower cutoff level at the first round and omission of the second round, resulting in an extended screening interval.

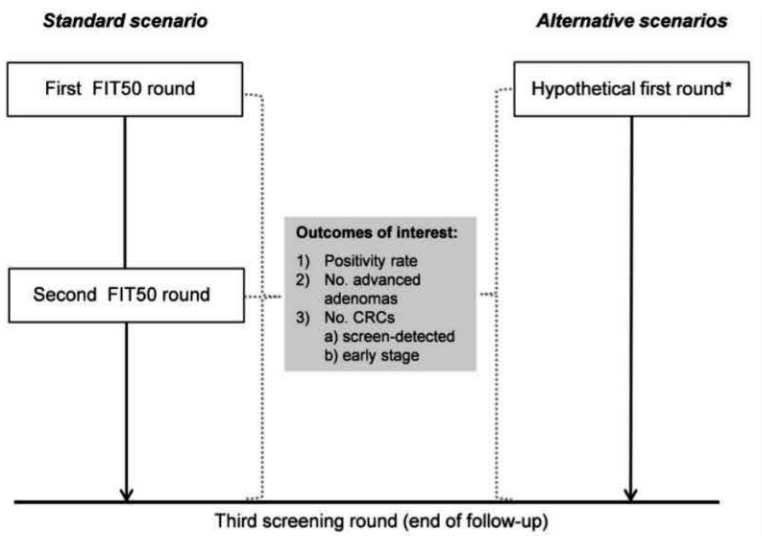


Figure 1. *Illustration of the concept of the data analysis and outcomes of interest.*

Study design and study population

Details about the design of the study have been described elsewhere [6,9]. Briefly, demographic data of all individuals aged 50-74 years living in the southwest of The Netherlands were obtained from municipal population registers to identify the target population. This population was screening-naïve since there was no CRC screening program at the time of recruitment for this study. Random samples were taken from the target population by a computer-generated algorithm (Tenalea, Amsteram, The Netherlands) to be invited for successive FIT screening rounds. A screening interval of 1, 2 or 3 years was assigned to equally sized groups between the first and the second round, while the screening interval was 2 years for all subjects between the second and the third round. Exclusion criteria were a history of CRC; inflammatory bowel disease; an estimated life-expectancy of less than 5 years; a colonoscopy, sigmoidoscopy or double-contrast barium enema within the previous

3 years; and inability to give informed consent. Subjects were no longer invited to subsequent rounds if they tested positive at a prior screening round, if they had become ≥75 years of age if they had moved out of the region or had died. The occurrence of CRCs in the study population was determined by record linkage with the Dutch Comprehensive Cancer Centre (http://www.iknl.nl). Results of the first, second and third round of the study have been reported previously[6,9,12].

Screening intervention

With each screening round, one FIT (OC-Sensor Micro, Eiken Chemical Co, Tokyo, Japan) was sent by mail to collect a single stool sample of one bowel movement. The test was considered positive when the hemoglobin (Hb) concentration in the FIT sample was ≥50 ng/ml, which corresponds to 10 µg Hb/g feces. Applying FIT at this cutoff level is called FIT50 in the following, and analogous terms are used for other cutoff levels. Subjects with a positive FIT50 were scheduled for colonoscopy within 4 weeks. Experienced endoscopists performed all colonoscopies. The maximum reach of the endoscope, adequacy of bowel preparation as well as the characteristics and location of any lesions were recorded. Experienced gastrointestinal pathologists evaluated all removed lesions. Patients with relevant findings entered a surveillance program according to the guidelines of the Dutch Society of Gastroenterology, while patients with a negative colonoscopy were considered not to require FIT screening for 10 years[13].

Data analysis

For the present analyses, we only included subjects who participated in the first screening round because we required baseline fecal hemoglobin for our approach. The cohort was followed from the first round (baseline), over the second round up to the time point of the third screening round, yielding a follow-up period of 3, 4, or 5 years depending on the interval assigned between the first and the second round[9]. In a first step, we considered the standard scenario with two FIT50 rounds and determined the cumulative positivity rate (defined as the proportion of participants with a positive test result), the cumulative number of subjects diagnosed with advanced adenomas, as well as the total number of CRCs (detected either as a result of screening or during follow-up as interval CRC). We determined the number of screen-detected CRCs and non-screen-detected CRCs (i.e. interval CRCs). We also determined the number of CRCs that were diagnosed at an early stage (UICC I or II), regardless of whether they were screen-detected or not. In a second step, we estimated these outcomes for various alternative scenarios which assumed that a cutoff level equal or lower than 50 ng/ml had been used for FIT at the first round (baseline) and the second round had been skipped. The various alternative scenarios only differed with respect to the cutoff level

used to classify baseline hemoglobin levels as positive or negative. Positivity rates of the various alternative scenarios could be assessed directly by determining the proportion of the cohort whose baseline hemoglobin levels were equal or above the respective cutoff level. The number of subjects diagnosed with advanced adenomas at the alternative scenarios could not be assessed directly given that only those with hemoglobin levels ≥50 ng/ml underwent colonoscopy at baseline. Therefore, we used an indirect approach to estimate this number as described in Table 1.

Table 1. *Approach to estimate the number of subjects diagnosed with advanced adenomas at the various alternative scenarios.*

A) Input parameters: Sensitivities for advanced adenomas at various cutoff levels according to the results of a diagnostic study that determined test performance characteristics of OC-Sensor Micro (Eiken Chemical Co, Tokyo, Japan)[14].

Cutoff level [ng hemoglobin / ml]*	Sensitivity for advanced adenomas
≥11	50%
≥14	47%
≥22	45%
≥36	38%
≥45	36%
≥50	35%

B) Based on these sensitivities, the number of subjects diagnosed with advanced adenomas at the various alternative scenarios was estimated in a two-step approach.

Step 1:	*Estimating the total number of subjects with advanced adenomas in the cohort at baseline:* Given a sensitivity of 35% for advanced adenomas at a cutoff level of ≥50 ng/ml, those subjects diagnosed with advanced adenomas in the first FIT50 round (N=126) were considered to make up for 35% of all subjects with advanced adenomas in the cohort at baseline. Dividing N=126 by 35%, it can thus be estimated that 360 subjects in the cohort had advanced adenomas at baseline
Step 2:	*Estimating the number of subjects diagnosed with advanced adenomas at the hypothetical first round:* This number (N=360) was multiplied with the sensitivity for advanced adenomas at the cutoff level of each respective scenario. For example, when using a cutoff level of 22 ng/ml at the first round, it was estimated that 162 subjects would have been diagnosed with advanced adenomas (N=360 multiplied with 45%).

The approach made use of the results of a diagnostic study that was conducted in a similar study population in the same Dutch region and determined test performance characteristics of OC-Sensor Micro (Eiken Chemical Co, Tokyo, Japan) at various cutoff levels[14]. In this study, all included subjects underwent both FIT and colonoscopy. To determine the number of screen-detected CRCs for the alternative scenarios, we considered persons who were diagnosed with CRC until the end of the follow-up period (screen-detected or interval CRC) and assessed whether their baseline hemoglobin level was equal or above the cutoff level of the respective scenario. We also determined the number of early stage CRCs (UICC I or II) for the alternative scenarios, taking into account uncertainty regarding stage progression as far as relevant. Accordingly, we distinguished between CRCs that were definitely or possibly detected at an early stage. The assumptions and criteria we used for classifying CRCs as screen-detected and early stage in the alternative scenarios are described in

detail in the Supplement. We also determined the number of FITs and follow-up colonoscopies for the standard scenario and the various alternative scenarios.

Results

Overall, 4,523 subjects (48% male) participated in the first FIT50 round. The mean age (standard deviation) was 60.5 (6.6) years. At the second round, 742 subjects (16%) were no longer eligible, and 354 subjects (8%) were eligible but did not participate. This left 3427 (76%) participants in the second FIT50 round (91% of those being eligible). A flow chart of the cohort from baseline to the end of the follow-up period is provided in the Supplement. A total of 36 CRCs were diagnosed in the cohort; 22 with first-round FIT50, 6 with second-round FIT50, and 8 during intervals. Figure 2 provides information on these CRCs, including whether they were screen-detected with FIT50 or interval CRCs, the respective baseline FIT level, the stage at diagnosis and the time between baseline and diagnosis.

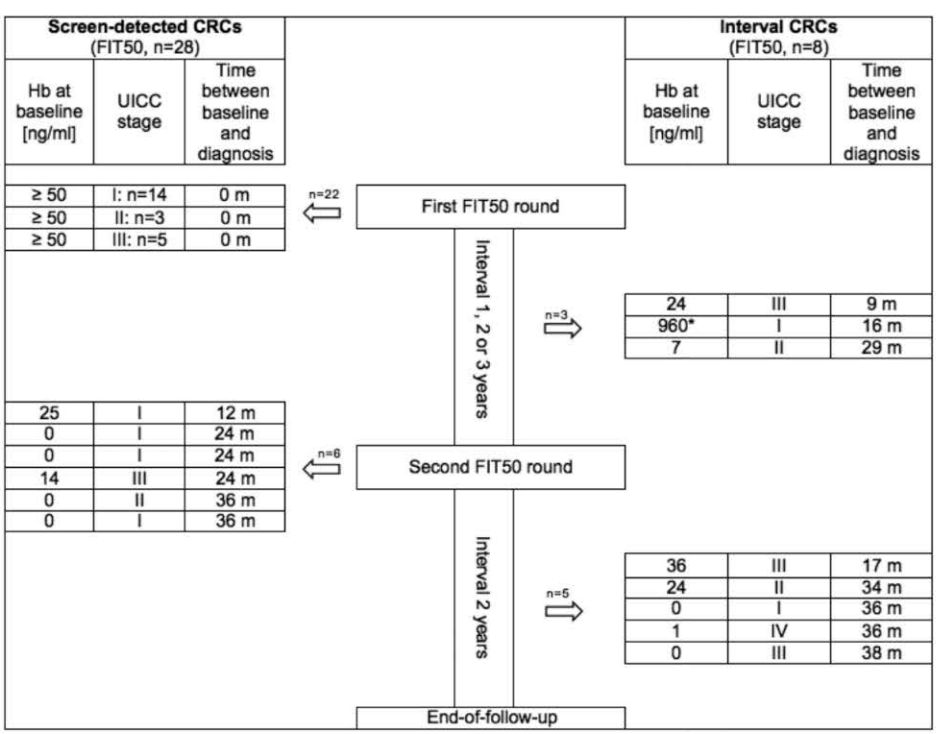


Figure 2. *Information on colorectal cancers (n=36) occurring in the cohort during the follow-up period, including the respective baseline FIT level, the stage at diagnosis and the time between baseline and diagnosis. * This person tested positive at the first FIT50 round but follow-up colonoscopy was negative.*

Table 2 shows the outcomes of interest for the standard scenario and the various alternative scenarios. In the standard scenario (i.e. the two FIT50 rounds), the cumulative positivity rate was 13%. Advanced adenomas were detected in 180 subjects, the majority of which (70%) were diagnosed in the first FIT50 round [9]. Overall, 26 CRCs (72% of all CRCs) were diagnosed at an early stage, of which 22 were screen-detected.

Regarding the alternative scenarios, Table 2 shows the results for selected cutoff levels ranging from 11 to 50 ng/ml. Using FIT11 for the alternative single initial screening round yielded a similar number of subjects with advanced adenomas as the standard scenario, while the positivity rate was 5% higher (i.e. 18 vs. 13%; $p<0.0001$). The number of CRCs diagnosed at an early stage was estimated to range between 22 and 27. The lower estimate (i.e. 22) only included CRCs that would definitely have been diagnosed at an early stage, while the upper estimate (i.e. 27) also included those that would possibly have been diagnosed at an early stage (see Methods section and Supplement). A single-round FIT22 scenario yielded a similar positivity rate as the two-round standard scenario, while the number of subjects diagnosed with advanced adenomas was estimated to be 10% lower. The number of early stage CRCs was estimated to range between 22 and 26. For FIT36, the positivity rate decreased to 10% and the number of subjects diagnosed with advanced adenomas was 25% lower as compared to the standard scenario. The number of early stage CRCs was estimated to range between 21 and 25. There was a trade-off between a higher number of FITs in the standard vs. a higher number of colonoscopies in the alternative scenarios when the latter used cutoff levels below 22 ng/ml. For alternative scenarios with cutoff levels ≥ 22 mg/ml both the number of FITs and colonoscopies were lower than in the standard scenario (Table 2).

Discussion

The quantitative nature of FIT as opposed to gFOBT offers new options for CRC screening that are yet to be fully explored. In this study, we explored FIT strategies that use a lower than usual positivity threshold in combination with an extended screening interval (up to 5 years), based on the analysis of data from repeated FIT screening. While such strategies would save screening rounds, our results suggest that they likely do not markedly differ from conventional strategies with respect to diagnostic yield and cumulative positivity rate, with some trade-off depending on the respective cutoff level. There were scenarios with similar lesion detection and a higher number of follow-up colonoscopies, and scenarios with slightly lower lesion detection and a similar number of follow-up colonoscopies. The number of FITs was, as a matter of course, halved compared to the standard

Table 2. *Positivity rate, number of subjects diagnosed with advanced adenomas, number of screen-detected colorectal cancers, number of early stage colorectal cancers and number of FITs and colonoscopies for the standard scenario and the respective estimates for various alternative scenarios.*

	Positivity rate	**Subjects with advanced adenomas* (n)**	**CRCs (overall: n=36)**				**FITs and colonoscopies (n)**	
			Screen-detected CRCs** (n)	Early stage CRCs (UICC I or II; n)				
				Definite	Possibly, category A3***	Possibly, category B3***	FITs	Colonoscopies
Standard scenario								
	13% (580/4523)	180	28	26	NA	NA	7950	580
Alternative scenario								
≥11	18% (826/4523)	180	27	22	3	2	4523	826
≥14	16% (745/4523)	169	27	22	3	2	4523	745
≥22	13% (578/4523)	162	26	22	2	2	4523	578
≥36	10% (460/4523)	137	23	21	1	3	4523	460
≥45	9% (410/4523)	128	22	21	0	3	4523	410
≥50	8% (380/4523)	126	22	21	0	3	4523	380

* For the hypothetical scenarios, the numbers of subjects with advanced adenomas were estimated as described in Table 1.

** The number of interval CRCs is calculated by subtracting the number of screen-detected CRCs from the overall number of CRCs (n=36).

***As described in more detail in the Supplement, these categories take into account that there is uncertainty regarding stage progression for some CRCs in the alternative scenarios. Category A: CRCs diagnosed at an advanced stage during the follow-up period that might have been at an early (or precancerous) stage if detected at baseline. Category B: This category refers to CRCs that were detected at an early stage at the second FIT50 round (i.e., the round that was skipped at the alternative scenarios) and had a hemoglobin level at the first round below the cutoff level of the respective alternative screening scenario. As justified in the Supplement, we considered half of these CRCs to be possibly still at an early stage if diagnosed later.

scenario due to the longer interval.

The next question is what motivates to further pursue this research, i.e. what may be advantages of such screening strategies. An obvious advantage is the reduction in the number of screening rounds, which saves efforts and costs related to the organization of each screening round and may also reduce the burden to the screenees. For example, offering screening from age 50 to 70 years requires 11 rounds when a 2-year interval is used, but only 5 rounds when a 5-year interval is used. Furthermore, the increased likelihood of detecting lesions at one single round going along with the higher per-test sensitivity (resulting from the lower positivity threshold) would be particularly advantageous for subjects who participate in screening on an irregular basis. For example, a person with advanced adenomas who participates in screening only once has a 50% chance of being detected by screening when FIT11 is used, but only a 35% chance when FIT50 is used. In this study, this potential advantage was not apparent because the proportion of subjects who participated in both FIT50 rounds was very high. However, it is expected to be relevant in settings with less favorable patterns of longitudinal adherence. This may occur when there is no or a suboptimal invitation system, as it is often the case in countries with a decentralized health system. For example, a retrospective cohort analysis from the United States evaluating adherence to repeated yearly FOBT showed that among 395.000 subjects who received exclusively FOBT and no other screening tool, about 40% were tested only once during a 5-year study period[15]. If strategies with a longer screening interval are used, the design needs to make sure that non-participants are re-invited after a reasonable time frame and not only several years later when invitation to the next screening round is due.

Colonoscopy capacity needs to be taken into account when discussing the implications of alternative FIT strategies that use a lower positivity threshold. Our results suggest that the total demand for work-up colonoscopy could be rather similar to conventional strategies, with some variation depending on the respective cutoff level. From a program perspective, the higher number of positive tests per invited birth cohort is more or less compensated for by the lower number of birth cohorts that is invited per round due to the longer interval. It is important to note that this compensation is achieved during the steady state of an established screening program. When a program based on such alternative FIT strategies is started, the time taken for complete roll-out may need to be adapted in settings with limited colonoscopy capacity.

To our knowledge, there are no other studies with a similar approach. We previously used microsimulation modeling to assess cost-effectiveness of various FIT screening strategies over a

period of 30 years, varying the cutoff level (50 - 200 ng/ml), the screening interval (1-3 years) and the age range to which screening is offered[16]. Screening at a cutoff level of 50 ng/ml was found to be more cost-effective than at higher cutoff levels, which held true for the range of explored intervals. This indirectly supports our finding that screening at a lower cutoff level in combination with a longer interval may not be disadvantageous. The current study adds to these findings by lowering the cutoff level together with extending the screening interval in ranges not previously explored.
Empirical evidence on FIT screening with intervals of three or more years is limited. The European guidelines for quality assurance in CRC screening and diagnosis recommend that the screening interval for FIT should not exceed three years, referring to three case-control studies from Japan[7,17-19]. These studies determined the risk of developing CRC and of dying from CRC according to FIT screening history However, the subgroup analyses that focused on the optimal screening interval had methodological issues regarding sample size and confounder adjustment as detailed in the Supplement. Apart from that, the positivity rate was 2.4% in these studies indicating that a high cutoff level was used. This is in contrast to our approach that compensated for the longer interval by lowering the cutoff level.

Experimental evidence for varying the screening interval between 1, 2, and 3 years without varying the cutoff level has recently been reported from the study that the present analysis was based on [9]. The detection rate of advanced neoplasia at the second round did not vary between groups assigned to the different intervals. Although the comparison was not powered to detect small differences in detection rate between groups, these findings question the intuitive thinking that less frequent FIT screening (at a constant cutoff level) inevitably decreases the cumulative diagnostic yield. With respect to our analysis, it suggests that the screening interval could be extended to 5 years when the skipped round is compensated by use of a lower cutoff level at the former round. Our approach used lower than usual positivity thresholds for FIT. Generally, this is a way to bridge the difference between fecal occult blood test screening and primary colonoscopy screening where all screenees are offered colonoscopy irrespective of any blood in stool. As shown in our analysis, the overall demand for colonoscopy of such strategies may be kept at the same level as for conventional FIT screening if the screening interval is extended. Diagnostic studies suggest that when lowering the positivity threshold to the hemoglobin levels that we used in our analysis, the increase in sensitivity relative to the decrease in specificity is similar to higher cutoff levels[14]. Accordingly, we expect similar conclusions if our approach is applied to data from programs that used higher cutoff levels. In other words, lowering the positivity threshold (although not to the levels we used in our analysis) and extending the interval could also be an option for settings with a lower colonoscopy capacity.

For some FIT products, analytical imprecision (that is, variability between measurements repeated under similar conditions) has been shown to vary according to the cutoff level[20]. Thus, if low cutoff levels are used for routine screening, the precision profiles need to be optimized accordingly. The recent optimization of buffers of several FITs has already contributed to this issue and allows use of lower cut-off levels.

There are strengths and limitations to this study that should be noted. We developed an approach that was directly linked to empirical longitudinal data of FIT screening. This avoided the need for assumptions regarding conditional independence of sequential FIT testing, which plays a key role in the context of exploring longer screening intervals. There were uncertainties regarding stage progression for some CRCs, which we tried to address in a systematic and transparent way. The same applies to the number of adenomas detected in the alternative scenarios, which we estimated by combining the findings of the first FIT50 round with the results of a study comparing FIT and colonoscopy in all subjects screened[14]. The latter was conducted in a similar setting and showed a similar detection rate for FIT50 regarding advanced adenomas (3.1% vs. 2.8% in the first FIT50 round). The sample size of the diagnostic study was moderate, but the course of the ROC curve at lower cutoff levels was confirmed by other diagnostic studies on quantitative FIT[21].
There is uncertainty whether the clinical benefit of detecting advanced adenomas that bleed less vs. detecting those that bleed more is similar. It may be that the latter are more likely to progress to lesions that would be symptom-detected at an early stage anyway (and would thus not require screen-detection), but it may also be that the level of bleeding is positively associated with the progressive potential. As possible limitations, it should also be noted that not all CRCs testing positive with FIT at baseline may actually have been detected at work-up colonoscopy and that de novo CRCs could have occurred that were not detectable at baseline.

Although imperfect, our approach can be considered as a step that helps deciding whether or not trials that directly compare conventional and alternative FIT strategies with an extended screening interval are justified. Further preliminary evidence from other databases on repeated FIT screening that are analysed in the same way would be of value: First, for the purpose of validation given the inherently low number of CRCs detected after the initial screening round in our study. Secondly, studies that used a higher cutoff level at the initial round would provide insights regarding generalizability to other settings. With respect to the time frame, our analysis was focused on skipping the FIT round that follows the initial screening. It will be interesting to conduct similar analyses for subsequent screening rounds to explore alternative FIT strategies from a longer-term

perspective. Finally, as the diagnostic yield of FIT differs by age and sex increased sample sizes would allow for subgroup analyses according to these factors.

In conclusion, our findings suggest that the diagnostic yield of alternative FIT strategies using a lower than usual positivity threshold in combination with an extended screening interval (up to 5 years) may be similar to conventional strategies. This justifies and motivates further research steps in this direction given that such alternative strategies could present interesting options for CRC screening, either generally or in particular settings.

References

1. Jemal A, Bray F, Center MM, et al. Global cancer statistics. CA Cancer J Clin 2011;61:69-90.
2. Hewitson P, Glasziou P, Irwig L, et al. Screening for colorectal cancer using the fecal occult blood test, Hemoccult. Cochrane Database Syst Rev 2007; CD001216.
3. Allison JE, Sakoda LC, Levin TR, et al. Screening for colorectal neoplasms with new fecal occult blood tests: update on performance characteristics. J Natl Cancer Inst 2007;99:1462-70.
4. Park DI, Ryu S, Kim YH, et al. Comparison of guaiac-based and quantitative immunochemical fecal occult blood testing in a population at average risk undergoing colorectal cancer screening. Am J Gastroenterol 2010;105:2017-25.
5. van Rossum LG, van Rijn AF, Laheij RJ, et al. Random comparison of guaiac and immunochemical fecal occult blood tests for colorectal cancer in a screening population. Gastroenterology 2008;135:82-90.
6. Hol L, van Leerdam ME, van Ballegooijen M, et al. Screening for colorectal cancer: randomised trial comparing guaiac-based and immunochemical faecal occult blood testing and flexible sigmoidoscopy. Gut 2010;59:62-8.
7. Segnan N, Patnick J, von Karsa L (eds.) (2010) European guidelines for quality assurance in colorectal cancer screening and diagnosis- first edition. European Commission, publications office of the European Union, Luxembourg.
8. Schreuders EH, Ruco A, Rabeneck L, et al. Colorectal cancer screening: a global overview of existing programmes. Gut 2015;64:1637-49.
9. van Roon AH, Goede SL, van Ballegooijen M, et al. Random comparison of repeated faecal immunochemical testing at different intervals for population-based colorectal cancer screening. Gut 2013;62:409-15.
10. Denters MJ, Deutekom M, Bossuyt PM, et al. Lower risk of advanced neoplasia among patients with a previous negative result from a fecal test for colorectal cancer. Gastroenterology 2012;142:497-504.
11. Crotta S, Segnan N, Paganin S, et al. High rate of advanced adenoma detection in 4 rounds of colorectal cancer screening with the fecal immunochemical test. Clin Gastroenterol Hepatol 2012;10:633-8.
12. Kapidzic A, Grobbee EJ, Hol L, et al. Attendance and yield over three rounds of population-based fecal immunochemical test screening. Am J Gastroenterol 2014;109:1257-64.
13. Nagengast FM, Kaandorp CJ; CBO Werkgroep. [Revised CBO guideline: Follow-up after polypectomy] (in Dutch). Ned Tijdschr Geneeskd 2001;145:2022e5.
14. de Wijkerslooth TR, Stoop EM, Bossuyt PM, et al. Immunochemical fecal occult blood testing is equally sensitive for proximal and distal advanced neoplasia. Am J Gastroenterol 2012;107:1570-8.
15. Gellad ZF, Stechuchak KM, Fisher DA, et al. Longitudinal adherence to fecal occult blood testing impacts colorectal cancer screening quality. Am J Gastroenterol 2011;106:1125-34.
16. Wilschut JA, Hol L, Dekker E, et al. Cost-effectiveness analysis of a quantitative immunochemical test for colorectal cancer screening. Gastroenterology 2011;141:1648-55.
17. Saito H, Soma Y, Koeda J, et al. Reduction in risk of mortality from colorectal cancer by fecal occult blood screening with immunochemical hemagglutination test. A case-control study. Int J Cancer 1995;61:465-9.
18. Saito H, Soma Y, Nakajima M, et al. A case-control study evaluating occult blood screening for colorectal cancer with hemoccult test and an immunochemical hemagglutination test. Oncol Rep 2000;7:815-9.
19. Nakajima M, Saito H, Soma Y, et al. Prevention of advanced colorectal cancer by screening using the immunochemical faecal occult blood test: a case-control study. Br J Cancer 2003;89:23-8.

20. Homepage of the World Endoscopy Organization: http://www.worldendo.org/assets/downloads/pdf/activities/fit_reports/gmec_fit_evaluation_report.pdf; accessed October 1st, 2015.

21. Haug U, Hundt S, Brenner H. Quantitative immunochemical fecal occult blood testing for colorectal adenoma detection: evaluation in the target population of screening and comparison with qualitative tests. Am J Gastroenterol 2010;105:682-90.

Supplementary files

Supplement 1

Assumptions and criteria for classifying CRCs as screen-detected CRCs and early stage CRCs in the alternative scenarios.

Screen-detected CRC:

A CRC in a person whose baseline FIT level was equal or beyond the cutoff level of the respective scenario was classified as screen-detected. This assumes that this CRC developed from a lesion that was detectable at baseline colonoscopy. The lesion may have been precancerous (i.e. not yet malignant) at baseline, but as this is not known we cautiously considered it as screen-detected rather than prevented CRC.

Early stage CRCs:

- CRC definitely detected at an early stage: This category includes CRCs that were diagnosed at an early stage at baseline or during the follow-up period and would have tested positive at the hypothetical single initial round (i.e. the baseline haemoglobin level was equal or above the cutoff level of the respective scenario). This category also includes interval CRCs diagnosed at an early stage.
- CRC possibly detected at an early stage, category A: This category includes CRCs that were diagnosed at an advanced stage during the follow-up period and would have tested positive at the hypothetical single initial round (i.e. the baseline haemoglobin level was equal or above the cutoff level of the respective scenario). If detected at baseline, they might have been at an early (or precancerous) stage.
- CRC possibly detected at an early stage, category B: This category refers to CRCs that were screen-detected at an early stage at the second FIT50 round of the standard scenario (i.e. the round that was skipped at the hypothetical scenarios), but would have tested negative at the hypothetical single initial round (i.e. the baseline haemoglobin level was below the cutoff level of the respective scenario). It is not plausible to assume that all of these CRCs would have been detected at an advanced stage at the alternative scenarios. They may: i) have remained at an early stage until being screen-detected at a following round or ii) have been symptom-detected at an early stage. The latter is expected for 40% of CRCs given the stage distributions observed in the pre-screening era and among interval CRCs.

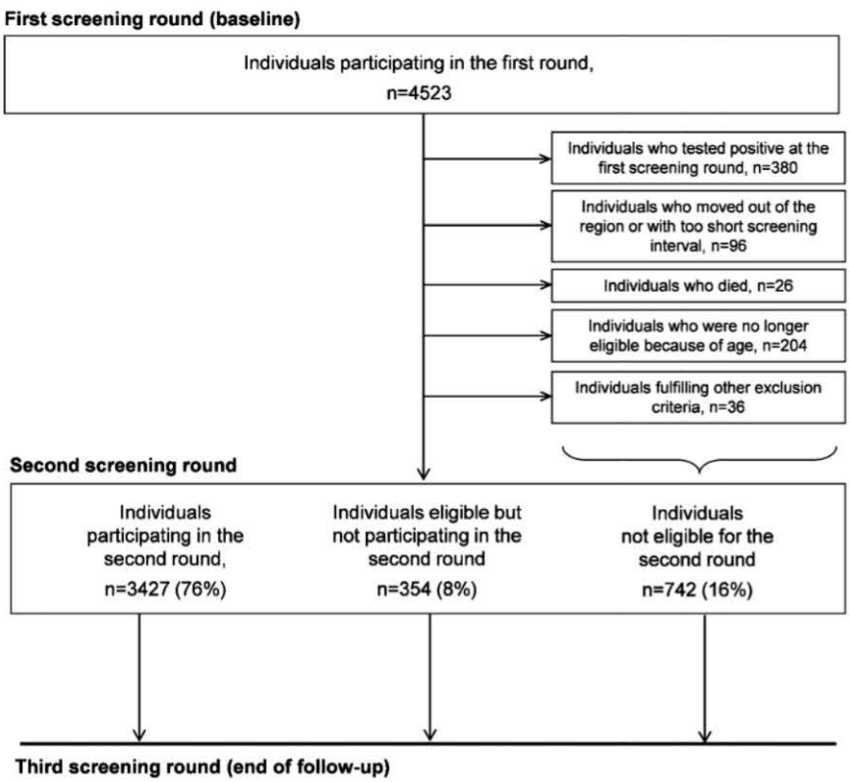
To take into account the possibilities described under i) and ii), we classified half of these CRCs as

being possibly detected at an early stage. For example, in the alternative scenario using a cutoff level of 14 ng/ml, 4 CRCs were in category B overall. We considered half of these CRCs (n=2) to possibly be still at an early stage if diagnosed later.

Supplement 2

Flow chart of the cohort from baseline to the end of the follow-up period.

It describes screening eligibility and participation at the second FIT50 round (corresponding to the round that would have been skipped at the alternative scenarios).



PART IV

quality and endoscopy in colorectal cancer screening


chapter
12

Systematic assessment of the quality of patient-oriented websites on colorectal cancer screening

Eline H. Schreuders, Esmée J. Grobbee, Ernst J. Kuipers, Manon C.W. Spaander, Sander J.O. Veldhuyzen van Zanten

Clinical gastroenterology and hepatology, 2016

Abstract

Background & Aims The efficacy of colorectal cancer (CRC) screening is dependent on participation and subsequent adherence to surveillance. The internet is increasingly used for health information and important to support decision-making. We evaluated the accuracy, quality, and readability of online information on CRC screening and surveillance.

Methods A Website Accuracy Score and Polyp Score were developed, which award points for various aspects of CRC screening and surveillance. Websites were also evaluated using validated internet quality instruments (Global Quality Score, LIDA and DISCERN), and reading scores. Two raters independently assessed the top-30 websites appearing in Google.comTM. Portals, duplicates and news articles were excluded.

Results Twenty websites were included. The mean Website Accuracy Score was 26 out of 44 (range 9-41). Websites with the highest scores were www.cancer.org, www.bowelcanceraustralia.org and www.uptodate.com. Median Polyp Score was three out of ten. The median Global Quality Score was three out of five (range 2-5). The median LIDA overall score was 74% and median DISCERN score was 45, both indicating moderate quality. The mean Flesch-Kincaid Grade Level was 11th grade rating the websites as difficult to read, 30% had a reading level acceptable for the general public (Flesch Reading Ease>60). There was no correlation between the Google rank and Website Accuracy Score (r_s=-0.31; p=0.18).

Conclusions There is marked variation in quality and readability of websites on CRC screening. Most websites do not address polyp surveillance. The poor correlation between quality and Google ranking suggests that screenees will miss out on high-quality websites using standard search strategies.

Introduction

Screening is effective in reducing the burden of colorectal cancer (CRC) and many countries have implemented CRC screening programs[1,2]. The success of CRC prevention is highly dependent on participation in the screening program. Initial participation and subsequent adherence to surveillance can be influenced by enhanced knowledge about CRC screening and colonoscopy outcome[3,4]. As more screening programs are implemented worldwide, providing adequate patient oriented information is increasingly important. Most organized screening programs approach individuals for screening on a voluntary basis without personal contact with a health professional[2,5]. Accordingly these individuals may search for additional information on screening themselves.

The internet is widely regarded as an important channel of health information[6,7]. In Western countries, more than half of the population uses a smartphone allowing instant and rapid access to the worldwide web[8]. However, few regulations control the information that individuals or organizations list on their web sites. A systematic review reported that 70% of studies identified quality issues with health- and disease-focused internet websites[9]. Since the efficacy of a CRC screening program is dependent on informed participation, assessing the availability and quality of online information aimed at screenees is of crucial importance. Therefore, the aim of this study was to rate quality, accuracy and readability of web-based information on CRC screening from a screenee perspective.

Methods

Internet Search Strategy

Web sites were identified by searching the World Wide Web with Google.comTM , the most frequently used Internet search engine[10]. The search was performed with English settings, with location tracking and search activity history switched off so search results were not influenced by location or past searches. Searches were carried out in 2014, 2015 and 2016. The following search terms were used: "colorectal cancer screening" OR "bowel cancer screening" OR "colon cancer screening" (quotations included). The search terms used reflect the most searched terms listed in the statistics provided via Google Trends. It is known that internet searchers do not typically view more than a few search hits and usually choose one of the first results displayed by the search engine[11]. We therefore decided to examine the first 30 hits, corresponding with the first three pages of Google searches.

Table 1. *Colorectal cancer screening specific Website Accuracy Score components and percentage of websites that were awarded points for these items.*

Website information components (maximum 44 points)	**Websites**
	N (%)
CRC general information	
Description of the colon/bowel/large intestine	15 (75)
Image of the anatomy of the intestines	15 (75)
Explanation of polyp as precursor of colorectal cancer	17 (85)
Development of a polyp into malignancy is a slow process (takes years)	8 (40)
Colorectal cancer can be prevented by removing precancerous polyps/adenomas	15 (75)
Risk factors	
Unknown	3 (15)
Age (>50 years)	13 (65)
Gender	0 (0)
History of previous polyps	15 (75)
Familiy history of colorectal cancer	17 (85)
Hereditary / Familial adenomatous polyposis / Lynch syndrome	14 (70)
Lifestlyle (2 points*):	
unhealthy life style (general)	
unhealthy diet (low fiber, high fat, red meat)	
smoking	
alcohol	
obesity	
* Mentions 1-2 life style factors: 1 point	1 (5)
Mentions 3 or more life style factors: 2 points	13 (65)
Symptoms of colonic polyps / CRC	
Most polyps are asymptomatic	11 (55)
Mentions symptom(s) such as:	13 (65)
blood in stool/ rectal bleeding	
change in bowel habit	
unexplained weight loss	
tenesmus (false urge)	
Recommendation to contact medical doctor in case of symptoms	11 (55)
Screening for CRC	
Mentions that there are different methods of screening	16 (80)
The detection and removal of polyps is main purpose of the screening program for colorectal cancer	15 (75)
Mentions that there are different methods of screening	16 (80)
The detection and removal of polyps is main purpose of the screening program for colorectal cancer	15 (75)
Mentions that not all tests have same accuracy	7 (35)

Mentions that not all tests have same patient burden	7 (35)
Colonoscopy is gold standard / most accurate for diagnosing polyps	7 (35)
Colonoscopy	20 (100)
+ explanation of procedure	17 (85)
+ explanation risks (bleeding and perforation are mentioned)	14 (70)
+ explanation polypectomy	13 (65)
+ explanation bowel preparation	13 (65)
Mentions flexible sigmoidoscopy	15 (75)
+ explanation procedure	13 (65)
+ explanation risks	8 (40)
Mentions FOBT (immunochemical or guaiac)	20 (100)
+ explanation procedure	16 (80)
+ has to be repeated every 1-2 years	13 (65)
+ stresses importance of repeated screening	7 (35)
+ explains possibility of false positive/negative results	11 (55)
Mentions barium enema	9 (45)
+ poor detection of (pre)cancer	3 (15)
Mentions CT colonography	11 (55)
+ explanation procedure	10 (50)
+ explanation risks	8 (40)
Mentions that all tests, when positive, need to be followed by colonoscopy	9 (45)
Mentions surveillance after colonoscopy in case of adenomas	3 (15)
Mentions that frequency of screening is different per test	7 (35)
Describes possibility of interval carcinomas (CRC after negative test)	4 (20)
Describes limitations of screening such as overdiagnosis and overtreatment	5 (25)

Inclusion and exclusion criteria

English websites were included only if the main part of the site dealt with educational information about CRC-screening. Websites that merely contained portal links to other sites were excluded, as were duplicate websites, news articles and sites containing irrelevant information (e.g. advertising, retail sites, or patient fora).

Accuracy assessment

The variability and accuracy of the information provided by each website on key facts about CRC screening and surveillance was investigated. For this purpose a Website Accuracy Score specific for CRC screening was developed (Table 1). In addition a separate Polyp Score for colorectal polyps was developed to assess information on important aspects of polyps, colonoscopy outcome and surveillance guidelines (Table 2). The Website Accuracy Score and Polyp Score consist of a list of key

items deemed relevant for CRC screening and surveillance. They were generated through evaluation of the literature and discussions with key stakeholders. The Website Accuracy Score and Polyp Score went through five iterations and were pretested twice prior to its final use using a random selection of websites. The range of scores was 0 to 44 for the Website Accuracy Score and 0 to 10 for the Polyp Score. If a website did not discuss or name an item of the Website Accuracy Score or Polyp Score, zero points were awarded for that item. Items had to be clearly presented on the website; the search function of the website was not used to locate this information.

Table 2. *Polyp Score items and percentage of websites that were awarded points for these items.*

Website information components	**Websites**
	N (%)
Polyp Score (maximum 10)	
Description of what a polyp is; growth/mushroom/lump in the lining of the large bowel	15 (75)
Image of a polyp	9 (45)
Prevalence of people with polyps in population	6 (30)
Explains that there are different types of polyps	9 (45)
Explains that not all polyps have an equal risk of turning in to colon cancer	10 (50)
Explains differences between adenoma and hyperplastic polyp	4 (20)
Mentions that some polyp characteristics have a higher risk of malignant degeneration; i.e. histological findings (villous aspect)	3 (15)
Polyp size as risk factor	3 (15)
Influence of degree of cleanliness of bowel on polyp detection	2 (10)
Explains surveillances intervals after polypectomy	2 (10)

Quality assessment

In addition to Website Accuracy Score and the Polyp Score, a selection of validated scores was used to assess the website quality and reliability. The overall quality of each website was rated using the Global Quality Score. This is a previously validated five-point Likert scale to rate the overall quality of a website (Table 3).12,13 It incorporates the accessibility of the information within the website, the quality of this information, the overall flow of information, and how useful the website reviewer thinks the particular website would be to a screenee. The Global Quality Score was assigned by the reviewer after evaluating the entire website.

The LIDA instrument is a validated question based instrument, assessing the overall score (0-96), accessibility (0-54), usability (0-12) and reliability (0-30) of healthcare websites. The scores are reported as percentages of the maximum score, overall scores >90% represent good results and <50% represent poor results. The online LIDA instrument was used for this study[14].

The DISCERN tool is a validated 16-item questionnaire to rate the quality of written information on treatment choices for a health problem.15,16_ENREF_18 The first 8 questions address reliability, dependability and trustworthiness of a website, the next 7 questions focus on quality of information on treatment choices and the last question addresses the overall quality of the site. Each question is rated on a 5-point scale with a maximum score of 80. Questions were answered as if participation to CRC screening was the treatment choice. The total quality of each website was classified as high (≥65 points), moderate (33–64 points), or low (16–32 points).

Table 3. *Global Quality Score criteria used to score websites on colorectal cancer screening.*

Score	Global Quality Score Description
1	Poor quality, poor flow of the site, most information missing, not at all useful for patients.
2	Generally poor quality and poor flow, some information listed but many important topics missing, of very limited use to patients.
3	Moderate quality, suboptimal flow, some important information is adequately discussed but other information poorly discussed, somewhat useful for patients.
4	Good quality and generally good flow, most of the relevant information is listed, but some topics nog covered, useful for patients.
5	Excellent quality and excellent flow, very useful for patients.

The amount of advertisements on each website was scored as none, little, average, or many and agreed through discussion by the two reviewers.

Readability assessment

Readability, referring to the reading difficulty based on word and sentence length, was assessed by the use of two readability scores. The Flesch Reading Ease Score (FRE) assigns a value between 0 and 100 whereby a higher value represents a greater ease of reading. A section with a score of 90–100 is considered to be very easily understood, >60 is an acceptable level of difficulty for the general public and below 30 is considered very difficult to read[17]. The Flesch-Kincaid Grade Level (FKG) uses the same input variables as the FRE score and outputs a US school grade indicating the average school grade able to read the text[17]. The American Medical Association Foundation states that health-related materials for patients should be written at a level appropriate for those in the 6th grade or below[18]. The FRE score and FKG score were calculated using the Microsoft Word 2007 program. A random 100 word sample of text was extracted from each website and pasted into the program by both reviewers independently.

Statistical analyses

The website assessment was performed by two independent raters (EHS, EJG). For Website Accuracy Score assessment, any difference in score between reviewers was resolved through discussion and by re-review of the website by both reviewers together to generate a single score for each website. Consensus in case of disagreement was achieved through discussion with a third reviewer (SvZ). For other quality parameters, the mean score of both website raters was used. Correlations between different quality parameters were analyzed using the Spearman rank-correlation coefficient because of non-normality of the data. Statistical tests were performed with the use of IBM SPSS software, version 21.0 and Graphpad Prism 5. A two-sided p-value <0.05 was considered significant. All authors had access to the study data and reviewed and approved the final manuscript.

Results

The search

The Google search was carried out on April 9th, 2014 and resulted in over 2,000,000 hits. The first 30 results were evaluated of which 20 websites were included. Two portal websites leading to another site, one duplicate site, one website with information on insurance reimbursement, three news articles, and three guidelines and medical articles clearly aimed at health professionals were excluded. All websites were accessed between April 2014 and June 2014. Additional Google searches were carried out on August 7th, 2015 and February 22nd, 2016 to evaluate possible changes in Google rank position. Most websites were published by a professional medical society (35%) or a governmental organization (30%) (Figure 1). Almost half of the websites were from the United States (45%), others from the United Kingdom (25%), Canada (20%) and Australia (10%).

Accuracy and quality of website information

The mean Website Accuracy Score was 26 (range 9-41). Most websites contained general information on CRC screening, but description and risk of different screening modalities and limitations of screening were not always captured (Table 2).

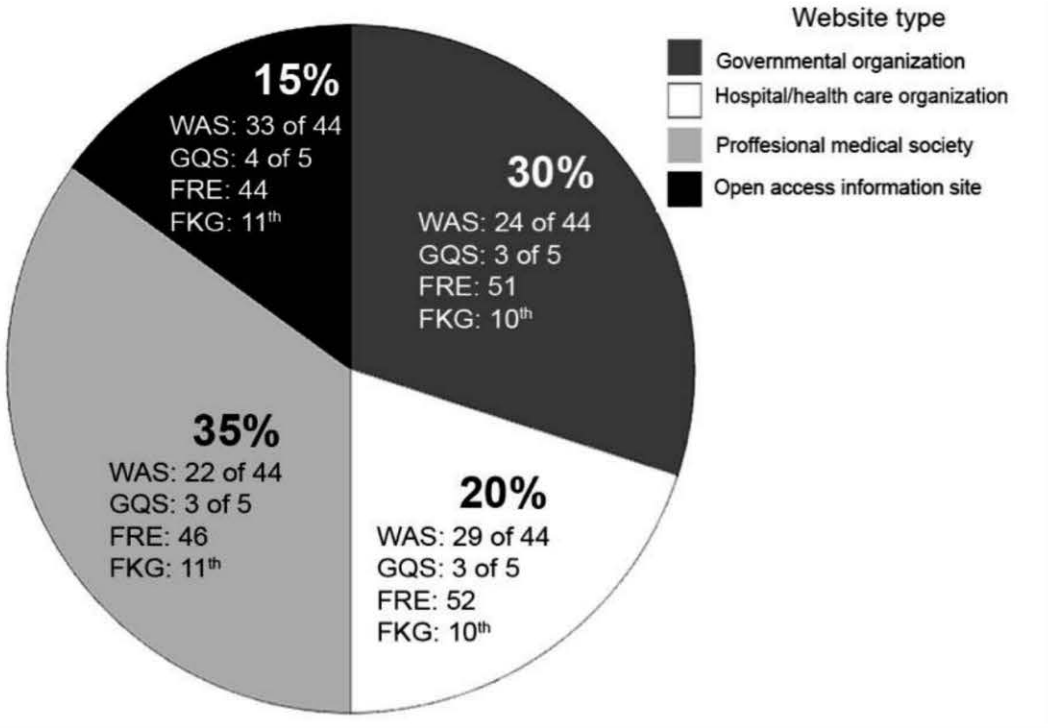


Figure 1. *Mean Website Accuracy Score, Global Quality Score and reading scores per website type.*

The median Global Quality Score was 3 (range 2-5). This score indicates that the quality of information of most websites was moderate. In many sites, some information was adequately discussed, while other parts of information were missing and the overall flow of information was suboptimal. There was a strong positive correlation between the Website Accuracy Score and the Global Quality Score with a Spearman's rho (rs) of 0.81 ($p<0.001$; Figure 2).

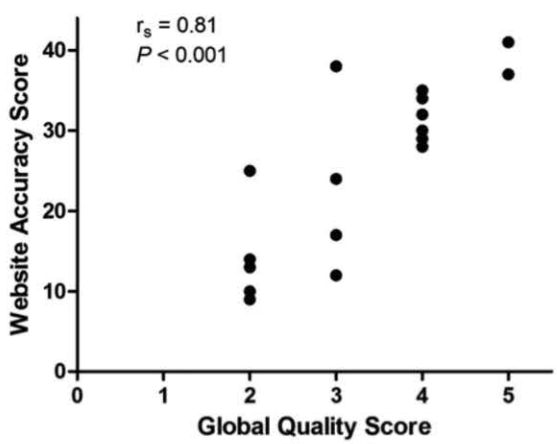


Figure 2. *Relationship of the Global Quality Score and the Website Accuracy Score used to evaluate colorectal cancer screening websites.*

The median Polyp Score was three (range 0-10, Table 2). The Polyp Score correlated positively with the Global Quality Score (rs=0.81; p<0.001). The median LIDA overall score was 74% (Interquartile range, IQR 11). The median LIDA score for accessibility was 88% (IQR 8), for usability 63% (IQR 22), and for reliability 52% (IQR 26). The median DISCERN score was 45 (IQR 20) indicating moderate quality. Ten percent of websites (2/20) were classified by DISCERN as high quality, 80% (16/20) as moderate and 10% (2/20) as low quality. Both the validated LIDA and DISCERN had a moderate correlation with the Website Accuracy Score; rs=0.45 (p<0.05) and rs=0.66 (p<0.01) respectively. There was no correlation between the Google ranks and the Website Accuracy Score (rs=-0.31; p=0.18, rs=-0.47; p=0.08 and rs=-0.31; p=0.25 for the 2014, 2015 and 2016 search respectively).

Table 4 lists the top five websites as rated by CRC screening specific Website Accuracy Score and other evaluations of website quality. The complete scores per website are published in a supplementary table. Eight websites had initial inter-rater Website Accuracy Score differences of ≥8. Differences in scoring of Website Accuracy Score or Polyp Score between reviewers were due to oversight or differences in interpretation.

Table 4. Top 5 websites as ranked by the Website Accuracy Score with the corresponding Polyp Score, quality scores, reading scores and Google rank positions.

Website	Accuracy		Quality			Readability		Google rank		
	WAS	PS	GQS	DISCERN	LIDA	FRE	FKG	2014	2015	2016
www.cancer.org	41	5	5	65	67%	62	9th	6	3	2
www.bowelcanceraustralia.org	38	2	3	35	58%	58	10th	5	X	29
www.uptodate.com	37	6	5	69	85%	28	14th	27	13	16
www.macmillan.org.uk	35	5	4	49	69%	48	11th	19	X	X
www.nlm.nih.gov/medlineplus	34	6	4	57	81%	59	8th	4	4	4

FRE; Flesch Reading Ease score, FKG; Flesch-Kincaid Grade Level, GQS; Global Quality Score, PS; Polyp Score, WAS; Website Accuracy Score, X; Not in the first 30 Google results.

Readability of websites

The mean Flesch Reading Ease Score was 48 (range 27–76), 30% of the websites had a reading level acceptable for the general public defined by a Flesch Reading Ease Score of >60. The mean Flesch-Kincaid Gradel Level was 11 (SD ±2.2, range 5–16), indicating that the text would be understandable to an average 11th grade US student. The reading level of healthcare and governmental websites was the easiest, whereas the reading level of open access information sites was the most difficult (Figure 2).

Advertisements

When assessing the amount of advertisement, 16 (80%) websites contained none, two (10%) contained a moderate amount of advertisement, and two (10%) websites contained many advertisements. The latter two were open access websites. Websites published by governmental organizations contained no advertisements.

Discussion

This study shows that there is marked variation in accuracy, quality and readability of information on CRC screening websites and that most websites do not address polyp surveillance. The best five websites as ranked by the Website Accuracy Score are www.cancer.org; www.bowelcanceraustralia.org; www.uptodate.com; www.macmillan.org.uk; and www.nlm.nih.gov/medlineplus. Their corresponding Google rank positions varied over time and some of these websites will be missed by standard Google searches (Table 4).

The poor correlation between website accuracy and Google ranking is especially concerning given the fact that Google is a prominent search engine.10 Internet users often do not go beyond the first page of a search, which can result in missing websites that provide high quality information. This problem has been identified before[12,13,19].

Even though surveillance after colonoscopy, especially if adenomatous polyps were found, is important for CRC screening to reach its maximal efficacy, it was only mentioned in 15% of the websites. Surveillance intervals are based on findings during colonoscopy[20]. However, clear and easy to understand information on how findings during a screening colonoscopy, i.e. adenomatous polyps, determine the follow-up surveillance recommendations was lacking in most sites. This is reflected in the low overall median Polyp Score (3 out of 10) and the fact that only two websites (10%) described the actual surveillance intervals. This is an important information gap since adherence to surveillance is influenced by enhanced knowledge[4]. Previous studies have shown that patients may not be sufficiently aware of important endoscopic findings and the consequences this has for subsequent surveillance recommendations[3,21].Understanding the need of surveillance likely will motivate participants to adhere to surveillance recommendations.

The reading difficulty of most websites was far above the required standard. Only 5% of the websites met the recommended level by the American Medical Association Foundation of 6th grade or

below[18]. This suggests that most websites are too difficult for the average reader and this may result in misunderstanding of information. Other studies evaluating patient information websites also documented that the required reading levels were high and above the recommended 6th grade level[22-24]. Our study showed that commercially funded websites were more difficult to read than governmental websites. This is in accordance with previous literature[22].

When evaluating the Website Accuracy Scores, it became apparent that most websites only focused on the predominant screening test used in the country where the website originated, and did not provide information on other options for CRC screening. It is debatable whether it is necessary to inform screenees about all possible screening tests that are available[25]. However, providing information that several different options exist may help individuals, who are interested in screening, to make an informed decision[25,26]. Colonoscopy and guaiac or immunochemical Fecal Occult Blood Test (FOBT) were described in all websites in detail. However, not all websites stressed the importance of the need for repeated screening when FOBT is used. This in spite of strong evidence that repeating stool testing at regular intervals is of paramount importance for FOBT-based screening to be effective in the long term[1]. Only 20% of the websites mentioned the possibility of the occurrence of interval carcinomas. This may in part be explained by the fact that this aspect of CRC screening has only gained a lot of attention during the last few years. However, not mentioning potential limitations of screening may stand in the way of informed decision-making[27].

A strength of this study is that the Website Accuracy Score and Polyp Score are CRC screening specific evaluation tools. These content specific outcome measures showed moderate to strong correlation with the validated generic outcome measures of Global Quality Score, LIDA and DISCERN. This provides further evidence that the use of these CRC screening specific outcome measures provide meaningful and relevant information. The advantage of the Global Quality Score over LIDA and DISCERN is that it is short and easy to perform. We believe that the Global Quality Score is a good score for overall flow and ease of use of any website providing health information.

This is to our knowledge the first review that systematically assessed the quality, accuracy and readability of patient-oriented websites on CRC screening as well as polyp surveillance. Previous studies have reported on the quality of web-based information regarding CRC surgery or treatment but none were systematic reviews of existing websites[23,28,29]. Two other publications evaluated CRC screening websites but these did not include detailed information on polyps and surveillance[30,31]. An American study focused on the readability and suitability of 12 CRC screening websites[31]. However,

these sites were self-chosen by the author. Another brief review examined five chosen websites and evaluated their content and usability[30]. In both of these publications no apparent selection criteria for quality were used. Most of the listed websites did not appear in our original 2014 search results, nor in the first three Google pages assessed in 2015 and 2016.

Our study has some limitations. Both the Website Accuracy Score and Polyp Score were not validated separately prior to use on the selected websites. However, the good correlation with other previously validated quality instruments suggests adequate content validity. We only searched using English search terms thus only English websites were retrieved. Another possible limitation is the fact that quotation based search terms were used which require words to appear together in retrieved websites. Omission of quotation marks when searches are done could lead to different results.

The internet is increasingly used by consumers to find relevant health information. There is evidence that experience and knowledge of Internet use has a significant impact on the uptake of CRC screening[32]. Furthermore, the credibility of cancer-related information on the internet is associated with population compliance with CRC screening, indicating the relevance of this study[4,32]. We believe health care providers interested in developing websites on CRC screening, for example for their own institutions, can use our approach to evaluate the quality and readability of provided information to develop the content of the site they are creating. Alternatively they can provide health care consumers to several of the high quality websites, listed in table 4 that we identified.

Physicians should be aware of the limitation of Google searching for CRC screening. Our study may be helpful in that regard as it provides a list of those websites which provide the highest quality information on CRC screening. However, it is important to remember that the internet is "alive" and that quality of websites may change over time or that new high quality websites may be developed.

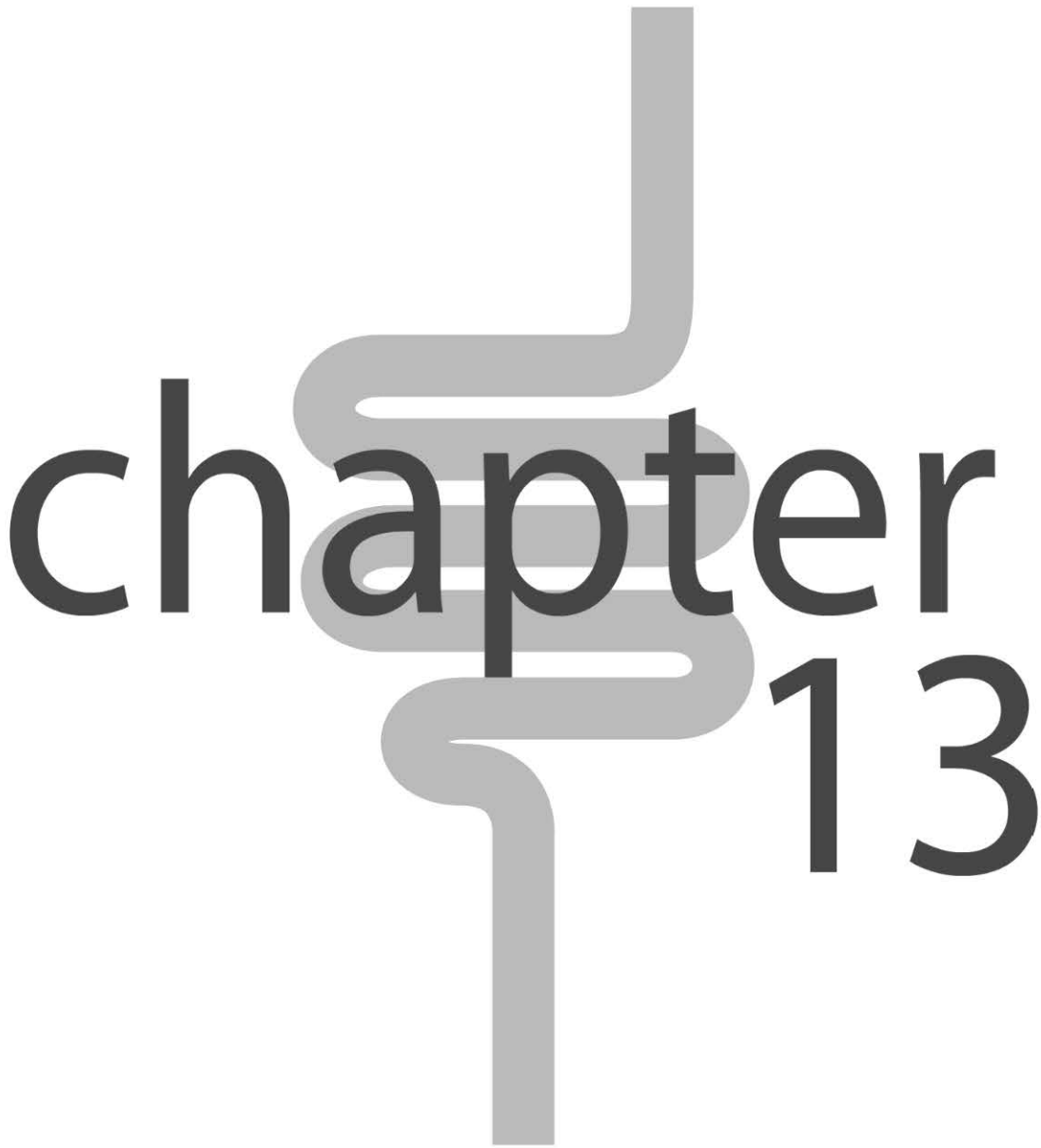
In conclusion, our study showed that there is marked variation in overall quality of web-based patient information on CRC screening. Most websites lack important information regarding polyps and their importance for future follow up surveillance colonoscopies. Several high quality websites do exist but poor correlation with Google ranking suggests that these websites may be missed. High quality and readable websites are essential to provide patients with reliable information to make informed decisions on CRC screening and surveillance participation and to optimize efficacy.

References

1. Kuipers EJ, Rosch T, Bretthauer M. Colorectal cancer screening--optimizing current strategies and new directions. Nat Rev Clin Oncol 2013; 10(3): 130-42.
2. Schreuders EH, Ruco A, Rabeneck L, et al. Colorectal cancer screening: a global overview of existing programmes. Gut 2015; 64(10): 1637-49.
3. Kumaravel V, Heald B, Lopez R, Hasson H, Schneider K, Burke CA. Patients Do Not Recall Important Details About Polyps, Required for Colorectal Cancer Prevention. Clinical Gastroenterology and Hepatology 2013; 11(5): 543-7.e2.
4. Wee CC, McCarthy EP, Phillips RS. Factors associated with colon cancer screening: the role of patient factors and physician counseling. Prev Med 2005; 41(1): 23-9.
5. 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: gutjnl-2014-308074.
6. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010; 12(1): e4.
7. Zulman DM, Kirch M, Zheng K, An LC. Trust in the internet as a health resource among older adults: analysis of data from a nationally representative survey. J Med Internet Res 2011; 13(1): e19.
8. eMarketer. Worldwide Smartphone Usage to Grow 25% in 2014. Jun 11, 2014 2014. http://www.emarketer.com/Article/Worldwide-Smartphone-Usage-Grow-25-2014/1010920.
9. Eysenbach G, Powell J, Kuss O, Sa ER. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA 2002; 287(20): 2691-700.
10. comScore. qSearch 2015 U.S. Desktop Search Engine Rankings. January 20, 2016 2016. https://www.comscore.com/dut/Insights/Rankings/comScore-Releases-December-2015-US-Desktop-Search-Engine-Rankings.
11. Eysenbach G, Kohler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. BMJ 2002; 324(7337): 573-7.
12. Bernard A, Langille M, Hughes S, Rose C, Leddin D, Veldhuyzen van Zanten S. A systematic review of patient inflammatory bowel disease information resources on the World Wide Web. Am J Gastroenterol 2007; 102(9): 2070-7.
13. Langille M, Bernard A, Rodgers C, Hughes S, Leddin D, van Zanten SV. Systematic review of the quality of patient information on the internet regarding inflammatory bowel disease treatments. Clin Gastroenterol Hepatol 2010; 8(4): 322-8.
14. Minervation. Minervation validation instrument for health care websites. http://lidaminervationcom/ 2011.
15. British Library and the University of Oxford. The DISCERN Instrument http://wwwdiscernorguk/discern_instrumentphp 2007; (22-12-2014).
16. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999; 53(2): 105-11.
17. Flesch R. A new readability yardstick. J Appl Psychol 1948; 32(3): 221-33.
18. Weis BD. Health Literacy and Patient Safety: Help Patients Understand: Manual for Clinicians: American Medical Association Foundation and American Medical Association; 2007.
19. Hargrave DR, Hargrave UA, Bouffet E. Quality of health information on the Internet in pediatric neuro-oncology. Neuro Oncol 2006; 8(2): 175-82.
20. Lieberman DA, Rex DK, Winawer SJ, et al. Guidelines for colonoscopy surveillance after screening and polypectomy: a

consensus update by the US Multi-Society Task Force on Colorectal Cancer. Gastroenterology 2012; 143(3): 844-57.

21. Sint Nicolaas J, de Jonge V, Cahen DL, et al. Awareness of surveillance recommendations among patients with colorectal adenomas. Clin Gastroenterol Hepatol 2012; 10(4): 405-11.
22. Risoldi Cochrane Z, Gregory P, Wilson A. Readability of consumer health information on the internet: a comparison of U.S. government-funded and commercially funded websites. J Health Commun 2012; 17(9): 1003-10.
23. Grewal P, Alagaratnam S. The quality and readability of colorectal cancer information on the internet. Int J Surg 2013; 11(5): 410-3.
24. Hutchinson N, Baird GL, Garg M. Examining the Reading Level of Internet Medical Information for Common Internal Medicine Diagnoses. Am J Med 2016; 129(6): 637-9.
25. van Dam L, Kuipers EJ, Steyerberg EW, van Leerdam ME, de Beaufort ID. The price of autonomy: should we offer individuals a choice of colorectal cancer screening strategies? Lancet Oncol 2013; 14(1): e38-46.
26. Hol L, de Bekker-Grob EW, van Dam L, et al. Preferences for colorectal cancer screening strategies: a discrete choice experiment. Br J Cancer 2010; 102(6): 972-80.
27. Biesecker BB, Schwartz MD, Marteau TM. Enhancing informed choice to undergo health screening: a systematic review. Am J Health Behav 2013; 37(3): 351-9.
28. Wasserman M, Baxter NN, Rosen B, Burnstein M, Halverson AL. Systematic review of internet patient information on colorectal cancer surgery. Dis Colon Rectum 2014; 57(1): 64-9.
29. Al-Bahrani A, Plusa S. The quality of patient-orientated internet information on colorectal cancer. Colorectal Dis 2004; 6(5): 323-6.
30. Saini SD. Website Review: Review of Patient-Oriented Websites for Colorectal Cancer Screening. Gastroenterology 2015; 148(3): 661-2.
31. Tian C, Champlin S, Mackert M, Lazard A, Agrawal D. Readability, suitability, and health content assessment of web-based patient education materials on colorectal cancer screening. Gastrointest Endosc 2014; 80(2): 284-90.
32. Chen CC, Yamada T, Smith J. An evaluation of healthcare information on the Internet: the case of colorectal cancer prevention. Int J Environ Res Public Health 2014; 11(1): 1058-75.


chapter
13

Screen-detected and non-screen-detected colorectal cancers after four rounds of fecal immunochemical test-based colorectal cancer screening.

Esmée J. Grobbee, Manon van der Vlugt*, Patrick M. Bossuyt, Amanda Bos, Evelien Bongers, Wolfert Spijker, Iris Lansdorp-Vogelaar, Ernst J. Kuipers, Manon C.W. Spaander, Evelien Dekker *both authors contributed equally*

Manuscript in preparation

Abstract

Objective Fecal immunochemical test (FIT)-based colorectal cancer (CRC) screening aims to detect CRC in an early stage, thereby reducing morbidity and mortality from this disease. Little information is available on the comparison of screen-detected CRC (SD-CRC) and non-screen-detected CRC (non-SD-CRC).

Design Between 2006 and 2014, asymptomatic persons aged 50 to 74 were invited to four biennial FIT-screening rounds. CRCs were identified through linkage with the Netherlands Cancer Registry and were classified into four groups: SD-CRC, FIT interval cancers, colonoscopy interval cancers and CRC in non-participants. The latter three categories represent non-SD-CRC. We compared patient characteristics, tumor site, stage and outcome between the four groups.

Results A total of 27,304 eligible individuals were invited for FIT-screening, of whom 18,716 (69%) participated at least once. Of these, 3,005 (16%) had a positive FIT in one of the 4 screening rounds. In total, 261 patients developed CRC within the four groups: 116 SD-CRC, 27 FIT interval CRC, 9 colonoscopy interval CRC, and 109 CRC were detected in non-participants. Patient characteristics did not differ significantly between groups. SD-CRC, FIT interval cancers and CRC in non-participants were mostly located in the distal colon (71%, 63%, 62%, respectively); while colonoscopy interval cancer were more often in the proximal colon (67%; p=0.01). Stage distribution differed between the four groups, with more favorable stages in patients with SD-CRC (p<0.001), and comparable distributions for FIT interval cancer and CRC in non-participants (p=0.39). Patients with SD-CRC and FIT interval cancers had a significantly higher survival rate than those with colonoscopy interval cancer and non-participants with CRC.

Conclusion In this FIT-screening program, participants with SD-CRC had more favorable stages and a better survival compared to those with interval CRC and to non-participants with CRC. Our results support the effectiveness of FIT-screening programs.

Introduction

Colorectal cancer (CRC) is one of the leading causes of cancer-related deaths in the Western World[1,2]. Approximately half of all diagnosed persons will die from the disease[3]. Survival is strongly related to tumour stage at time of diagnosis, with a 5-year survival of 94% for stage I CRC to 8% for stage IV CRC[3]. Population-based screening for CRC aims to detect cancer in an early stage to reduce CRC-related mortality.

High participation rates, together with the performance characteristics of a screening test, are crucial for screening effectiveness. While colonoscopy is considered the reference standard for detecting CRC and advanced adenomas, participation in primary colonoscopy screening is generally low [4,5]. For this reason, screening with fecal occult blood test (FOBT) as a first-line test is typically preferred, as participation rates in FOBT screening are considerably higher[1,5-7]. Essentially, two types of FOBT screening exist; guaiac FOBT (gFOBT) screening and fecal immunochemical test (FIT) screening. The latter is gaining popularity, as it has several advantages, such as ease of use and better diagnostic accuracy[8].

In contrast with screen-detected CRC (SD-CRC), a proportion of screening invitees are diagnosed with CRC outside a screening program: so-called non screen-detected CRC (non-SD-CRC)[9]. A part of these non-SD-CRC are interval cancers (IC), which are defined as cancers detected after a negative screening exam and before the date of the next recommended screening[9]. These cancers can be missed by either the primary first-line screening test (e.g. FIT IC) or they develop after a negative colonoscopy, following a positive FIT test (colonoscopy interval cancer within a FIT screening program)[9]. The remaining part of non-SD-CRC are among those who did not participate in screening.

Monitoring the incidence of non-SD-CRC, especially IC, is a crucial part of the evaluation of any CRC screening program, and is an indicator of program sensitivity. Previous studies involving gFOBT screening have shown better survival rates of participants with screen-detected CRC, compared to those with interval cancers[10]. Only very limited data exist on cancers in FIT-based screening programs, following participants and non-participants over a longer follow up period.
The aim of our study was to compare patient demographics, tumor site, stage and survival between patients with SD-CRC and non-SD-CRC in a population-based FIT screening program.

Methods

Population and design

Since 2006, two pilot programs of biennial FIT-based CRC screening have been conducted in the southwest and northwest regions of the Netherlands. Those two cohorts were combined for a fourth round of screening in 2014. Details about the design of these CRC screening programs have been reported previously[11,12]. In short, demographic data of all invitees between 50 and 74 years living in the target areas were obtained from municipal population registers. For the southwest region, random samples were taken from the target population by a computer-generated algorithm (Tenalea, Amsterdam, The Netherlands). In the northwest region random samples of selected postal code areas were taken. No national screening program had been implemented at the start of this pilot-program, and thus the target population was screening-naive when first contacted. In the Netherlands, a national FIT-based CRC screening program was gradually initiated from January 2014 onwards. Invitees for our cohort were not invited for the national program. Selected persons were invited for each consecutive round, except for those who had moved out of the area, those that had passed the upper age limit, institutionalized people, invitees unable to give informed consent, and those who had tested positive in a previous screening round. In our information leaflet, persons with a history of inflammatory bowel disease or CRC were asked not to participate CRC screening. Participants reporting a colonoscopy in the past 2 years during intake after a positive FIT were excluded from further participation and those with an estimated life expectancy of less than 5 years.

Characteristics

Date of birth, sex, and postal codes of all invitees were collected from the municipal population register. Socioeconomic status was based on the "social status scores" that are available through the Netherlands Institute of Social Research (www.scp.nl). The social status score of a postal code area was derived from the unemployment rate, educational level, average income and position on the labor market. The standard deviation (SD) in the Netherlands in 2006 (start of pilot) was 0.96. Socioeconomic status was divided into three categories based on this SD into high, average and low SES. The first available postal code of the invitee was used to categorize invitees.

Invitations and stool tests

Since 2006, a FIT and an invitation letter were sent to invitees by postal mail biennially. Due to organizational and logistic issues the time interval has fluctuated between rounds (median time between invitations was 2.37 years; IQR 2.01 to 2.76). Invitations were sent between June 2006 and

December 2014 (first round June 2006-February 2007; second round August 2008-June 2009; third round February 2011-February 2012, fourth round March 2014-December 2014). The FIT could be returned by (free) postal mail. Two different brands of stool tests were used in our dynamic cohort. In the first, second, and third round all invitees received an OC-sensor (Eiken Chemical Co, Tokyo, Japan). In the fourth round (executed between March 2014 and February 2015), all invitees were randomly allocated to receive either an OC-Sensor or an FOB Gold (Sentinel Diagnostics SpA, Milan, Italy).

Fecal Immunochemical Test

A single FIT was sent to the screenee. Returned FITs were handled by one of two specialized laboratories. A hemoglobin value of 10 µg Hb/g feces was used as the positivity threshold; this corresponds to 50 ng/ml buffer for OC-Sensor and 58 ng/ml buffer for FOB-Gold. Actual FIT levels were recorded for all participants, in each round.

Test results

Screenees were informed about their test result by postal mail within 2 weeks.

Screenees with a negative test result received a letter explaining that no blood had been detected in the stool sample and no follow-up was needed at that time. It was emphasized that the FIT is not 100% sensitive and that vigilance for symptoms of CRC remained important. They were instructed to contact their general practitioner in case of symptoms, despite of the negative test result.

Screenees with a positive test result were invited for a consultation at the outpatient clinic to discuss the test result and follow-up. In the absence of contra-indications and after informed consent, a colonoscopy was scheduled, within 2 weeks of the consultation. A colonoscopy was not advised if the colon had been imaged two years ago or less (colonoscopy or CT-colonography), in case of a life expectancy of less than 5 years, or severe co-morbidity.

Follow-up colonoscopy and colorectal lesions

The colonoscopy was performed according to the international quality guidelines and all quality parameters were collected in a database [13]. Advice regarding surveillance colonoscopy after removal of adenomatous polyps, large (≥10 mm) serrated lesions or cancer was given to the client according to the Dutch CBO consensus[14]. Screenees with a negative colonoscopy were not re-invited from the population screening program until 10 years later. Data on the location, size, macroscopic aspects, morphology, as well as details on the technique for polypectomy and endoscopic assessment of radicality were recorded for all colorectal lesions detected during colonoscopy.

Collected lesions were evaluated by an experienced gastrointestinal pathologist, using the Vienna criteria[15]. Cancers were staged according to the 7th edition of the American Joint Committee on Cancer classification[16].

Screen detected cancers, non-screen detected cancers and reference population

After finishing the fourth round in December 2014, the cohort was linked to the Dutch Comprehensive Cancer Centre for data between May 2006 and March 2015. Data on tumor stage, location and survival were collected for all CRC cases. All persons identified with CRC were categorized into groups as defined by Sanduleanu et al[9] (Table 1). The proportion of FIT interval cancer was calculated by dividing the number of FIT interval cancers by the sum of SD-CRC and FIT interval cancers. Persons with CRC detected at a scheduled surveillance colonoscopy and CRC occurring in patients with a positive FIT who did not undergo a subsequent colonoscopy (due to refusal), are reported separately. The median time between invitations (2.37 years; IQR 2.01-2.76) was used as a cut-off to categorize patients within the FIT interval cancer category. If a patient developed a CRC within this interval, after a negative FIT, but was not (yet) invited for a consecutive round (having passed the upper age limit, or moved out of the area), the patient was categorized as having a FIT interval cancer. As a reference group we used all individuals diagnosed with CRC in the Netherlands during the same time period, and in the same age range (50-76 years), in that part of the population which had not been offered CRC screening. The upper range of 76 years was chosen based on the median time between invitations and the upper age limit for eligibility (74 + 2.37 years). The data on our reference group and on non-participants (since there was no informed consent) were anonymously analyzed and delivered by the Dutch Comprehensive Cancer Centre. Data on follow-up time, in days after diagnosis, and SES were not available for these groups.

Data-analysis

Tumor location was categorized as either distal to the splenic flexure or proximal (splenic flexure and proximal). Median follow up time was calculated from date of diagnosis to death or considered censored at the end of follow-up (31 March 2015). Separately, survival curves were estimated for patients with SD-CRC, FIT interval cancers, and post-colonoscopy cancers. Differences in proportions between groups were evaluated for statistical significance using the $\chi 2$-test statistic. The log-rank test statistic was used comparing survival curves. P-values <0.05 were considered statistically significant differences. Data analysis was performed using SPSS 23 for Windows (Chicago, Ill).

Ethics approval

Ethical approval for the study was provided by the Dutch National Health Council (WBO 2642467, 2832758, 3049078 and 161536-112008, The Hague, The Netherlands).

Table 1. *Definitions of cancers.*

Terminology	Definition
SD-CRC	*Screen-detected colorectal cancer* defined as CRCs detected at colonoscopy following a positive FIT result
non-SD-CRC	*Non screen-detected colorectal cancer* defined as all CRCs that were not diagnosed at a colonoscopy following a positive FIT result an divided into three groups: 1. FIT interval cancer defined as cancers diagnosed between screening rounds after negative FIT before the next FIT was due 2. Colonoscopy interval cancer defined as cancers diagnosed after negative colonoscopy after a positive FIT within the surveillance interval 3. Colorectal cancer in non-participants defined as cancers diagnosed in those who never took part in CRC screening
Reference	*The reference group* was defined as all patients diagnosed with CRC in the Netherlands population during the same time period and in the same age range (50-76 years), that were not offered CRC screening.

Results

Patient demographics and colorectal cancers

A total of 27,304 people were eligible for FIT-screening, of whom 18,716 (69%) participated at least once; 8,588 (31%) never participated (Figure 1). Among participants, 3,005 (16%) had a positive FIT in one of the 4 screening rounds. Colonoscopy was performed in 2,762 of these (92% adherence). Over four rounds of population-based FIT screening, CRC was diagnosed in 160 (0,9%) participants, with 73% being SD-CRC, 24% being FIT interval cancer or colonoscopy interval cancer, 3% being CRCs in FIT positive persons not adhering to colonoscopy and 2% being surveillance detected CRCs. Among non-participants, 109 (1.3%) CRCs were diagnosed.

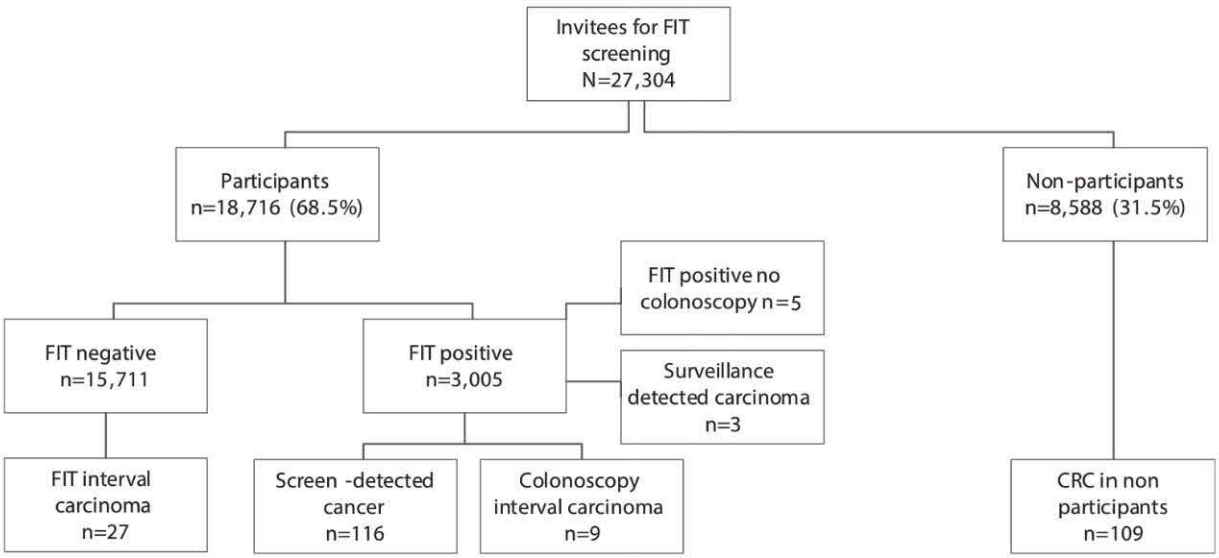


Figure 1. *Flowchart of the 4 rounds of FIT-based CRC screening including screen-detected and non-screen-detected colorectal cancers among participants (participating at least once) and non-participants (never participated).*

In our cohort, 269 patients were diagnosed with CRC, of which 116 (43%) were SD-CRC and 145 were non-SD-CRC. The latter group was made up by 27 (10%) FIT interval CRC, 9 (3.3%) colonoscopy interval CRC, and 109 (41%) CRC detected in non-participants. Three CRCs were detected at scheduled surveillance (1.1%) and five (1.9%) in patients not undergoing a colonoscopy after a positive FIT. These eight tumors were excluded from our further analyses. The FIT interval proportion was 19% (FIT interval cancers divided by the sum of FIT interval cancer and SD-CRC).

Table 2.*Characteristics of patients with screen-detected CRC, non-screen-detected CRC and the reference group.*

	Screen-detected CRC	**Non-screen-detected CRC**				**Reference group**
	SD-CRC	FIT interval cancer	Colonoscopy interval cancer	CRC in non-participant	p-value	
	(n=116)	(n=27)	(n=9)	(n=109*)		n=72612
Age at diagnosis (%)					0.831	
50-59	28 (24)	6 (22)	1 (11)	21 (19)		14651 (20)
60-69	50 (43)	11 (41)	6 (67)	49 (45)		31868 (44)
>70	38 (33)	10 (37)	2 (22)	39 (36)		26093 (36)
Sex (%)					0.709	
Male	73 (63)	16 (59)	4 (44)	69 (63)		42486 (59)
Female	43 (37)	11 (41)	5 (56)	40 (37)		30126 (42)
SES score (%)					0.763	
Low	13 (11)	2 (7)	2 (22)	N.A.		N.A
Average	82 (71)	21 (78)	6 (67)			
High	21 (18)	4 (15)	1 (11)			

*4 casus with 2 incidents, age at diagnosis first incident selected

Table 2 summarizes the characteristics of SD-CRC and non-SD-CRC separately. No significant differences were found regarding age at diagnosis between screening participants with SD-CRC and persons diagnosed with non-SD-CRC (p=0.83); most patients were diagnosed within the age group of 60-69 years. No gender differences were seen between the groups (p=0.71) with higher CRC rates in men in all groups : 63% males in the SD-CRC group, 59% in the group with FIT interval cancers, and 63% in the non-participants with CRC. Though more women than men were diagnosed with a post-colonoscopy interval cancer (44% versus 56%), this difference was not significant. There were also no significant differences in SES among the groups (p=0.76). The age and sex distribution was comparable with that of the reference population (no tested for statistical significance).

Tumor location and stage distribution

Tumor location and CRC stage distribution are described in Table 3. Screen-detected CRC, FIT interval cancers and CRC in non-participants were mostly located in the distal colon (71%, 63%, 62%, respectively), whereas colonoscopy interval cancers were more often located in the proximal colon (67%; p=0.063).

Table 3. *Tumor location and stage distribution of CRC in patients with screen-detected CRC, non-screen-detected CRC and the reference group.*

	Screen-detected CRC	**Non-screen-detected CRC**				**Reference group**
	SD-CRC	FIT interval cancer	Colonoscopy interval cancer	CRC in non-participant	p-value	
	(n=116)	(n=27)	(n=9)	(n=109*)		n=72612
Tumor location*						
Proximal	34 (29)	10 (37)	7 (78)	38 (35)		23976 (33)
Distal	82 (71)	17 (63)	2 (22)	67 (62)	0.063	47290 (65)
Unknown	0 (0)	0 (0)	0 (0)	4 (4)		1346 (2)
Stage						
I	60 (52)	8 (30)	2 (22)	17 (16)		14002 (19)
II	16 (14)	6 (22)	0 (0)	31 (29)	<0.001	18384 (25)
III	37 (32)	9 (33)	1 (11)	36 (33)		21819 (30)
IV	3 (3)	4 (15)	6 (67)	23 (21)		16909 (23)
Unknown	(0) (0)	0 (0)	0 (0)	2 (2)		(1498 (2)

* proximal (=splenic flexure and proximal)

The stage distribution differed significantly between the four groups, with more favorable stages in patients with SD-CRC (p<0.001). Stage distribution was similar for patients with FIT interval CRC and for non-participants with CRC (p=0.39). There was a statistically significant difference between FIT interval cancers and SD-CRC (p=0.019).

All-cause mortality and 3-years survival

All-cause mortality rates were significantly lower for patients with SD-CRC and those with FIT interval cancers compared to non-participants, and patients with colonoscopy interval CRC (p<0.001; Table 4).

Table 4. *All-cause mortality of patients with screen-detected CRC, non-screen-detected CRC and the reference group.*

	Screen-detected CRC	**Non-screen-detected CRC**				**Reference group**
	SD-CRC	FIT interval cancer	Colonoscopy interval cancer	CRC in non-participant	p-value	
	(n=116)	(n=27)	(n=9)	(n=109*)		n=72612
All-cause mortality	13 (11)	5 (19)	5 (56)	44 (40)	<0.001	25221 (35)
Follow-up (months, IQR)*	50 (25-76)	40 (15-63)	19 (9-30)	-		-

* follow-up after the diagnosis of CRC

Survival in the reference group of CRC cases was most similar to survival observed in non-participants. Median time after diagnosis for SD-CRC, FIT-interval cancers and for colonoscopy interval cancers was 46.1 months (IQR 18,1-72,1). No follow-up time could be calculated for our reference group or for non-participants, since only year of diagnosis was provided and not the exact date of diagnosis. There was a significant difference in survival between SD-CRC, FIT interval cancers, and colonoscopy interval cancers with a worse outcome for colonoscopy interval cancers (p<0.001; Figure 3).

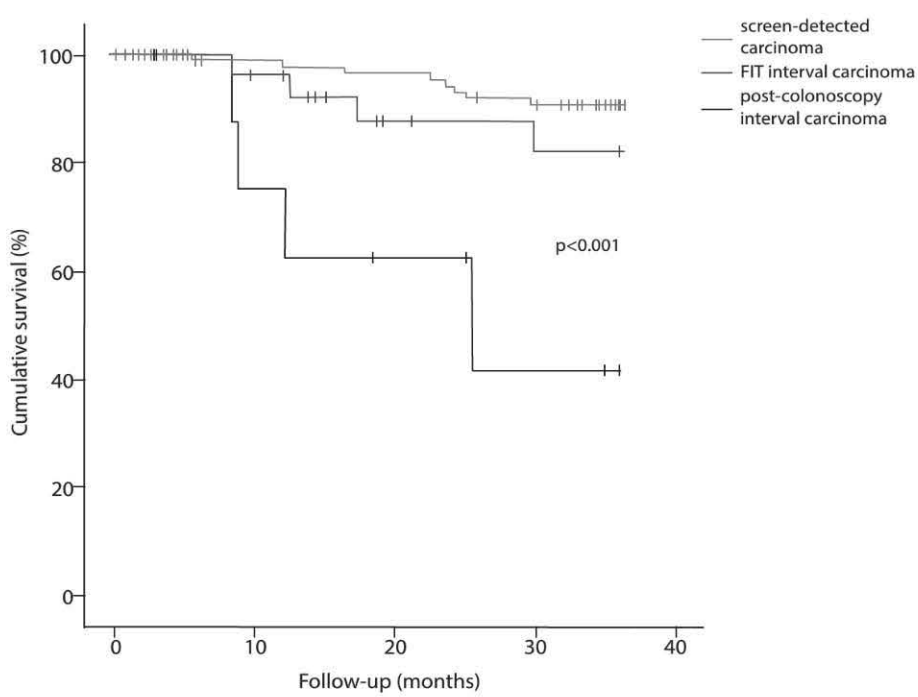


Figure 3. *Cumulative survival of screen-detected carcinomas, FIT interval carcinomas and post-colonoscopy interval carcinoma.*

Discussion

The results of this study support the effectiveness of FIT-based screening program as SD-CRC were diagnosed in an earlier stage than non-SD-CRCs. Furthermore, FIT interval cancer proportion was low when using a low cut-off in FIT-based CRC screening with better survival than the population that was not offered screening.

We are the first to describe interval cancers in a FIT-based screening program with long-term follow-up. Our cohort consists of an average-risk population, comprising all age ranges commonly invited for CRC screening programs worldwide. This population was screen-naive when first approached, without the presence of any other CRC screening initiatives in the population. All data were prospectively collected and all invitees were linked to the Netherlands Cancer Registry by the Dutch Comprehensive Cancer Centre to identify all non-SD-CRC. Since 1989, the Netherlands Cancer Registry registers all patients diagnosed with cancer in the Netherlands and provides a unique and fully covered database.

To fully appreciate our findings, some limitations also need to be addressed. Little information was available about non-participants developing CRC in our cohort because these data had to be delivered anonymously for privacy reasons. Their socioeconomic status could not be assessed. Previous studies have shown that especially socioeconomically deprived persons are less likely to take part in CRC screening though they presumably are at a higher risk of developing CRC due to poor general health[17,18]. Also the date of diagnosis could not be related to the date of invitation. It is possible that a person was already diagnosed with CRC prior to being invited for screening. In such case, refraining from participation would be a justified decision.

Although gFOBT screening has been shown to reduce mortality rates[19], high proportions of gFOBT interval cancers have been reported, ranging from 48% to 55%[20-23]. Since FIT is more sensitive in detecting advanced neoplasia than gFOBT, one would expect lower interval rates for FIT. A previous Scottish FIT-based study, however, reported a FIT interval proportion of 51% using a cut-off of 80 µg Hb/g feces[24]. This FIT interval proportion was calculated as the number of FIT interval cancers divided by the sum of FIT interval cancer and SD-CRC. We found a much lower proportion of FIT interval cancers (19%), which could be explained by the substantially lower positivity cut-off (10 µg Hb/g feces) in this cohort. Zorzi et al. reported a comparable FIT interval cancer proportion of 15% using a cut-off of 20 µg Hb/g feces[25]. The Scottish study investigated the effect of different

cut-off levels of fecal-Hb on interval cancer proportions, and colonoscopy demand It reported an interval cancer proportion of 51% at a cut-off of 80 µg Hb/g feces versus a proportion of 38% using a cut-off of 10 µg Hb/g feces with a significant increase in number of required colonoscopies for the latter cut-off[24]. Hence, choosing a very low positivity cut-off leads to a higher number of SD-CRC and lower proportion of FIT interval cancer but at the cost of increased colonoscopy demand and increased numbers false positives[24].

Our results showed no evidence for gender differences in interval cancer rates, while previous studies have shown higher rates of interval cancers among women[22,25-27]. As in other studies, women in our cohort were less likely to be diagnosed with CRC, but women were more often diagnosed with a post-colonoscopy interval cancer than men (44% versus 56%). Though actual numbers were small, and this difference did not reach significance. Our results are in line with a Spanish study, reporting SD-CRC and interval cancers during four rounds of FOBT screening using mainly gFOBT and a small portion FIT [23]. Digby et al. also did not observe a significant difference in gender for FIT, but suggested a trend towards higher rates of FIT interval cancers in women[24,26].
Previous gFOBT-based studies have shown an increase in the detection of earlier stages of CRC in screening, resulting in an overall improvement in survival for individuals diagnosed with SD-CRC compared to non-SD-CRC[10,19,20,28]. Our study showed that SD-CRC were most likely to be stage I, with a high survival rate, compared to non-SD-CRC, which confirms the findings of previous studies[10,23,26,29]. Next, in our FIT-based program a higher percentage of stage I CRC were found than in previously reported gFOBT screening programs. This could be possibly be explained by the use of a relatively low cut-off for FIT. Comparable early stage distributions in FIT screening have been reported[24,28]. Regarding location of the cancers, most tumors were located distally, except for colonoscopy interval cancers, which were mainly located in the proximal colon. These findings are in line with previous studies[23,30]. FIT does not seem to be less sensitive for FIT interval cancers in our cohort.

As SD-CRC are often asymptomatic and detected at an earlier stage, it could be expected that the outcome in these patients is better than patients with symptomatically diagnosed CRC (i.e. the reference group). This is known as the lead-time effect[31]. In contrast to previous literature which reported that prognosis of non-participants with CRC is actually poorer than that of patients with symptomatically diagnosed CRC, we observed that patient demographics, tumor location, tumor stage distribution, and survival rates of cancers arising were comparable between non-participants and the reference group[11]. Survival was significantly better in participants with FIT interval cancers than in non-participants with CRC. This is in contrast with findings of an English study that reported

similar outcomes for gFOBT interval cancers and a control group[22]. A possible explanation for this difference in outcome is the use of FIT (rather than gFOBT) and the use of a low cut-off. The better survival of FIT interval cancers compared to non-participants and the reference group, implicates that the theoretical risk of persons seeking delayed medical consultation in case of the development of abdominal symptoms due to (false) reassurance caused by a recent negative FIT result, fortunately, seems to have no immediate negative effect on all-cause mortality within our cohort. Nevertheless, these findings should be interpreted with caution since in theory FIT-participants could have been in better health than the reference group resulting in a selection bias.

In our cohort, nine colonoscopy interval cancers were detected. As a proportion (0.32%), this is comparable to a recently published FIT-based study, reporting a rate of 0.31%[29]. In those nine persons, the performed colonoscopies were reported as having been complete with cecal intubation. In 6 cases sufficient bowel preparation was reported, in 3 cases this data was missing. Three patients received a colonoscopy surveillance advice but developed a CRC within this interval. One person received the advice of a 1 year surveillance interval, 1 person a 3 year interval and 1 person a 6 years interval. Six patients were discharged from screening for 10 years as per protocol. As most colonoscopy interval cancers were detected in the proximal colon (78%), we should consider procedural factors, especially missed lesions (for instance due to inadequate bowel preparation). As both FIT interval cancers but also colonoscopy interval cancers within a FIT screening program are important for the quality and success of a screening program, both should be carefully monitored. Since the implementation of the national FIT-based CRC program in the Netherlands, a specialized clinical IT system has been developed to register, monitor and audit all quality indicator for colonoscopy according to the international quality guidelines[13]. All endoscopists performing endoscopies for our CRC screening need to be accredited. All these measures have been made to further improve quality of bowel cancer screening in order to reduce the number of post colonoscopy interval cancers and complications.

Five patients who did not want to undergo a colonoscopy after a positive FIT developed a CRC. Unfortunately, no information about their reason for refusal is available. To reduce these CRCs, additional strategies should be developed, to inform patients about the need for follow-up after a positive FIT. Modifiable determinants for non-adherence to endoscopy, such as embarrassment, lack of knowledge about CRC, fear of the procedure, or inconvenience should be discussed. Other strategies could be built on sending all non-respondents after a positive FIT result an additional information leaflet, or to offer them a consultation by phone.

In conclusion, our results show that survival in patients with FIT interval cancers in a population-based CRC screening program is better than in clinical CRC patients outside a screening program. For non-participants with CRC survival is more comparable to that of the reference population. Among all CRCs, post-colonoscopy interval cancers had the worst outcome, which stresses the need for insuring best quality colonoscopy according to the most recent quality guidelines. Overall, our results support the effectiveness of FIT-screening programs.

References

1. Ferlay J, Parkin DM, Steliarova-Foucher E. Estimates of cancer incidence and mortality in Europe in 2008. Eur. J. Cancer 2010;46:765–781.
2. Siegel et al. Colorectal Cancer Statistics 2014CA Cancer J Clin. 2014;64:104-117
3. Dutch Cancer Registry. www.cijfersoverkanker.nl
4. Stoop EM, de Haan MC, de Wijkerslooth TR, Bossuyt P, van Ballegooijen, Nio CY, van de Vijver MJ, Biermann K, Thomeer M, van Leerdam ME, Fockens P, Stoker J, Kuipers EJ, Dekker E. Participation and yield of colonoscopy versus non-cathartic CT colonography in population-based screening for colorectal cancer: a randomised controlled trial. Lancet Oncol 2012 Jan;13:55-64
5. Quintero E, Castells A, Bujanda L et al. Colonoscopy versus fecal immunochemical testing in colorectal-cancer screening. N Engl J Med. 2012;366:697-706.
6. Denters MJ et al. Lower risk of advanced neoplasia among patients with a previous negative result from a fecal test for colorectal cancer. Gastroenterol 2012;14:497-504.
7. Jensen CD, Corley DA, Quinn VP et al. Fecal Immunochemical Test Program Performance Over 4 Rounds of Annual Screening: A Retrospective Cohort Study. Ann Intern Med. 2016;;164:456-463.
8. 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 Aug;64(8):1327-1337.
9. Sanduleanu S, le Clercq CM, Dekker E, Meijer GA, Rabeneck L, Rutter MD, Valori R, Young GP, Schoen RE. Definition and taxonomy of interval colorectal cancers: a proposal for standardising nomenclature. Gut 2014 Sep 5. pii: gutjnl-2014-307992.
10. Gill MD, Bramble MG, Hull MA, Mills SJ, Morris E, Bradburn DM, Bury Y, Parker CE, Lee TJ, Rees CJ. Screen-detected colorectal cancers are associated with an improved outcome compared with stage-matched interval cancers. Br J Cancer. 2014;111:2076-2081.
11. Stegeman I, van Doorn SC, Mundt MW, Mallant-Hent RC, Bongers E, Elferink MA, Fockens P, Stroobants AK, Bossuyt PM, Dekker E. Participation, yield, and interval carcinomas in three rounds of biennial FIT-based colorectal cancer screening. Cancer Epidemiol. 2015;39:388-393.
12. Kapidzic A, Grobbee EJ, Hol L et al. Attendance and yield over three rounds of population-based fecal immunochemical test screening. Am J Gastroenterol 2014;109:1257-1264.
13. Rex DK. Quality in colonoscopy: cecal intubation first, then what? Am J Gastroenterol 2006;101:732-734.
14. Nederlandse richtlijn coloscopie surveillance. NVMDL nov 2013.
15. Schlemper RJ, Riddell RH, Kato Y, Borchard F, Cooper HS, Dawsey SM, Dixon MF, Fenoglio-Preiser CM, Fléjou JF, Geboes K, Hattori T, Hirota T, Itabashi M, Iwafuchi M, Iwashita A, Kim YI, Kirchner T, Klimpfinger M, Koike M, Lauwers GY, Lewin KJ, Oberhuber G, Offner F, Price AB, Rubio CA, Shimizu M, Shimoda T, Sipponen P, Solcia E, Stolte M, Watanabe H, Yamabe H. The Vienna classification of gastrointestinal epithelial neoplasia. Gut 2000;47:251-255.
16. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. Ann Surg Oncol. 2010;17:1471-1474.
17. Pornet C, Dejardin O, Morlais F et al. Socioeconomic determinants for compliance to colorectal cancer screening. A multilevel analysis. J Epidemiol Community Health 2010;64:318-324.
18. Moss SM, Campbell C, Melia J. et al Performance measures in three rounds of the English bowel cancer screening pilot. Gut 2012;61:101-107.

19. 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. N Engl J Med 1993;328:1365-1371.
20. Hardcastle JD, Chamverlain JO, Robinson MH, Moss SM, Amar SS, Balfour TW, et al. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. Lancet 1996;348:1472-1477.
21. Kronborg O, Fenger C, Olsen J, Jorgensen OD, Sondergaard O. Randomised study of screening for colorectal cancer with faecal-occult-blood test. Lancet 1996;348:1467-1471.
22. Gill MD, Bramble MG, Rees CJ. et all. Comparison of screen-detected and interval colorectal cancers in the Bowel Cancer Screening Programme. Br J Cancer2012;107:417-421.
23. Garcia et al. Interval Cancers in a Population-Based Screening Program for Colorectal Cancer in Catalonia, Spain. Gastroenterology Research and Practice Volume 2015:1-9.Digby J, Fraser C, Carey F et al. Interval ancers using a quantitative faecal immunochemical test (FIT) for haemoglobin when colonoscopy capacity is limited. J Med Screen 2015;0:1-5.
24. Zorzi M, Fedato C, Grazzini G, et al. High sensitivity of five colorectal screening programmes with faecal immunochemical test in the Veneto Region, Italy. Gut. 2011;60:944-949.
25. Steele RJC et al. Interval cancers in a FOBT-based colorectal cancer population screening programme: implications for stage, gender and tumour site. Gut 2012;61:576-581.
26. Brenner H, Chang-Claude J, Seiler CM, et al. Interval cancers after negative colonoscopy: population-based case-control study. Gut 2012;61:15768-15762.
27. Pande R, Froggatt P, Baragwanath. Survival outcome of patients with screening versus symptomatically detected colorectal cancers. Colorect Dis 2012;15:74-79.
28. Mansouri et al. A comparison of tumour and host prognostic factors in screen-detected versus non screen-detected colorectal cancer: a contemporaneous study. Colorectal Dis. 2016 Feb 9.
29. le Clercq CM, Sanduleanu S. Interval colorectal cancers: what and why. Curr Gastroenterol Rep. 2014 Mar;16(3):375.
30. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Early diagnosis. In: Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical epidemiology: a basic science for clinical medicine. 2d ed. Boston: Little, Brown, 1991:153–70.


chapter
14

Prevalence and treatment of T1 colorectal carcinoma in a FIT-based colorectal cancer screening program

Els Wieten, Esmée J. Grobbee, Paul Didden, Arjun D. Koch, Ernst J. Kuipers, Marco J. Bruno, Manon C.W. Spaander

Submitted

Abstract

Background and study aims The worldwide implementation of colorectal cancer (CRC) screening programs has led to an increased detection of early CRC. There is limited literature available on the prevalence of early CRCs in average CRC screening populations that can be curably resected by endoscopy (T1). We assessed the prevalence of patients with histological proven T1 colorectal carcinomas (pT1 CRC) in our screening cohort and how they were treated in current practice. Additionally, we assessed if histopathologic criteria for complete endoscopic resection were met according to the Dutch guidelines.

Patients and methods A random sample of 17,614 persons from the general Dutch population, aged 50-74 years, were invited to participate in a FIT-based CRC screening program. Participants with a positive test were referred for colonoscopy. In accordance with the guidelines, T1 CRCs were considered amiable for endoscopic resection in case the tumour was good or moderately differentiated and there was no lymphovascular invasion present. Complete R0 resection was achieved if all resection margins were ≥1 mm free from cancer.

Results CRC was detected in 86 subjects, of which 28 (33%) were diagnosed with pT1 CRC. Eighteen (64%) individuals with T1 CRC received endoscopic resection as initial treatment and ten (36%) were directly referred for surgery. In addition, nine patients underwent a surgical operation subsequent to endoscopic resection. Tumour morphology was significantly different between T1 CRCs initially treated by endoscopy (pedunculated) and those that were directly referred for a surgical operation (sessile), p=0.02. Of all 28 T1 CRCs, histopathology showed poor differentiation grade in one (4%) T1 CRC, lymphovascular invasion in two (7%). Endoscopic R0 resection was achieved in 33% (6/18). Histopathologic criteria for curable endoscopic resection were met in 28% (5/18).

Conclusions A substantial proportion of CRCs detected in a FIT-based CRC screening is a pT1 CRC and, according to the guidelines, amiable for endoscopic resection. In current practice, the majority of pT1 CRCs is eventually treated by surgical resection due to various reasons. The endoscopic approach to CRC should balance surgical morbidity against the potential benefits of endoscopic resection. Effort should be made to improve awareness of screening endoscopists to completely resect T1 CRCs.

Introduction

The worldwide implementation of colorectal cancer (CRC) screening programs has led to an increased detection of CRC in early stages of the disease[1,2]. Early-stage CRCs (T1) are observed more often in patients who participate in a CRC screening program than in symptomatic patients[3-5]. Although screening endoscopists are increasingly confronted with T1 colorectal carcinomas, the visual distinction at colonoscopy between these early CRCs and non-malignant adenomas remains challenging [6].

A radical endoscopic resection is a curative treatment for patients with an early CRC that is limited to the upper third of the submucosa (T1) and with no additional histopathological risk factors for lymph node metastasis[7]. In case unfavourable histopathologic features, such as poor differentiation grade of the tumour and/or lymphovascular invasion, are present in the tumour, additional surgery is still required[8,9]. However, information concerning these histopathological features is lacking at the time of the colonoscopy procedure. Therefore, intention to perform a radical endoscopic resection of the lesion is needed and could prevent patients from additional surgery.

Endoscopic resections by endoscopic submucosal dissection (ESD) and endoscopic mucosal resection (EMR) have emerged as curative and minimal invasive alternatives to surgery in selected cases with T1 colorectal carcinomas. Minimization of invasive procedures is of crucial importance in screening programs, as screening involves healthy subjects. Though, safety of the endoscopic procedure is another important outcome parameter since endoscopic resections have been associated with risks of serious complications and recurrence[10]. Therefore, the endoscopic approach to invasive T1 colorectal carcinomas must balance surgical morbidity against the potential benefits of endoscopic resection. International guidelines recommend to consider ESD for endoscopic removal of T1 carcinomas, particular for those larger than 20 mm[11,12]. ESD enables endoscopists to achieve en bloc resection regardless of the tumour size whereas EMR is only deemed suitable for en bloc resections of lesions ≤20 mm. However, ESD is technically demanding and associated with higher complication rates than EMR[13,14].

There is limited literature available on the prevalence of T1 colorectal carcinomas in average CRC screening populations that can be adequately resected by endoscopy according to guidelines and how these lesions are treated in current practice. Our CRC screening program consists of a large cohort with data from four rounds of biennial faecal immunochemical test (FIT)-based screening.

First, we assessed the prevalence of patients with histological proven T1 colorectal carcinomas (pT1 CRC) in our cohort. Secondly, we assessed the endoscopic aspects of these cancers and how they were treated. Thirdly, we estimated if patient characteristics or tumour aspects were associated with the initial choice of endoscopic or surgical treatment. Finally, we assessed if histopathologic criteria for complete endoscopic resection were met in these pT1 CRCs according to the guidelines.

Methods

Study design and population

Data was available from four biennial rounds of faecal immunochemical test (FIT)-based CRC screening in a formerly screening-naïve Dutch population. Details of the study have been published before[15,16]. In total, 17,614 subjects (aged 50 to 74 years) were randomly invited once or more between November 2006 and October 2014. In each screening round invitees were asked to perform one- or two-sample FIT (OC-Sensor Micro, Eiken Chemical, Japan).[15,16] Participants were referred for colonoscopy in case of at least one positive test using a cut-off of ≥ 50 ng/ml, which corresponds to ≥10µg haemoglobin (Hb) per gram faeces. Exclusion criteria were a history of inflammatory bowel disease or CRC, an estimated life expectancy of less than five years, and if the individual underwent a colonoscopy, sigmoidoscopy or barium contrast enema in the last three years.[15] The original studies were approved by the Dutch Ministry of Health (PG/ZP 2.727.071, PG/ZP 2.823.158). All screenees gave written informed consent. For the purpose of the present study, all patients diagnosed with a pT1 CRC were included, based on the pathology reports of the resected surgical and endoscopic specimen. Interval carcinomas were not taken into account. Data were retrieved from medical records, colonoscopy-, and pathology reports to assess tumour size, morphology, location, histopathologic features, and endoscopic resection techniques used.

Treatment guidelines and definitions

Dutch guidelines were followed for treatment recommendations. These guidelines are comparable to those of the American Society for Gastrointestinal Endoscopy (ASGE).[17,18] Definitions were based on the 2011 TNM (Tumour Node Metastasis) classification and Vienna classification.[19,20]

T1 CRCs were defined as CRCs confined to the submucosa, without invasion of the muscularis propria or deeper wall.[21] CRC in situ, which comprises a lesion restricted to the epithelial layer, was not considered as T1 CRC, but as non-invasive high-grade dysplasia (stage 0).

In case an R0 resection was performed in a T1 CRC with favourable histologic features, the tumour was considered curatively resected by endoscopy. Complete R0 resection was achieved if all resection

margins were ≥1 mm free from cancer[17]. Unfavourable histopathologic features of malignant colonic polyps included, poorly differentiated histology and/or lymphovascular invasion. According to the guidelines additional surgery is warranted if unfavourable histopathologic features are present in the tumour and/or resection margins were not free of tumour. Morphology of the pT1 CRC was described according to the Paris classification: pedunculated, sessile, slightly elevated, flat, slightly depressed or excavated[22]. Location of the pT1 CRC was considered proximal or distal according to the location with respect to the splenic flexure.

Statistical analysis

Comparisons of continuous variables were performed using the Mann–Whitney U-test. Categorical variables with two or more categories were compared by using the $\chi 2$ test. All p-values were two-sided and considered significant if $p<0.05$.

Results

In total, 9,327 eligible subjects participated at least once in four rounds of a one- or two-sample biennial screening. From the 2,280 participants that had a positive FIT in one of the four rounds, 2,117 (93%) underwent colonoscopy. Overall, colorectal cancer was detected in 86 subjects, of which 28 (33%) patients were diagnosed with a pT1 CRC.

Table 1. *Initial treatment displayed per patient characteristics and endoscopic aspects of T1 carcinomas.*

	Initial endoscopic resection	**Initial surgical resection**	**p-value**
	(n=18)	(n=10)	
Patient characteristics			
Sex, male (n, %)	14 (78)	6 (60)	
Age, > 60 years	12 (67)	8 (80)	0.45
Endoscopic aspects			
Tumour location proximal (n, %)	2 (11)	4 (40)	0.74
Tumour size ≥ 20 mm (n,%)	3 (17)	5 (50)	0.06
Tumour morphology (n, %)			
Sessile	3 (17)	5 (50)	
Pedunculated	10 (56)	0	
Flat	1 (6)	0	0.02
Unknown	4 (22)	5 (50)	
Total adenomas per patient > 2 (n, %)	7 (39)	4 (40)	0.95

sEndoscopic aspects at initial colonoscopy of pT1 colorectal carcinomas

Polyp morphology was described as pedunculated in ten (36%) T1 CRCs, sessile in eight (29%), flat in one (4%) and in nine (32%) T1 CRCs no polyp morphology was reported. Mean size as measured by the endoscopists of all T1 CRCs was 16 mm (range 4-40 mm). Median number of adenomas in patients with a T1 CRC was two, ranging from one to 49 adenomas. This latter patient with 49 adenomas was diagnosed with familial adenomatous polyposis (FAP). Of all 28 pT1 CRCs, seventeen (61%) were located in the sigmoid, four (14%) were located in the rectum, four (14%) in the ascending colon, two (7%) in the cecum , and one (4%) in the descending colon.

Resection of pT1 colorectal carcinomas

In total, nineteen (68%) of all 28 patients with a pT1 CRC had eventually undergone a surgical operation. Eighteen (64%) patients were initially treated by endoscopy and ten (36%) were directly referred for surgery (Figure 1).

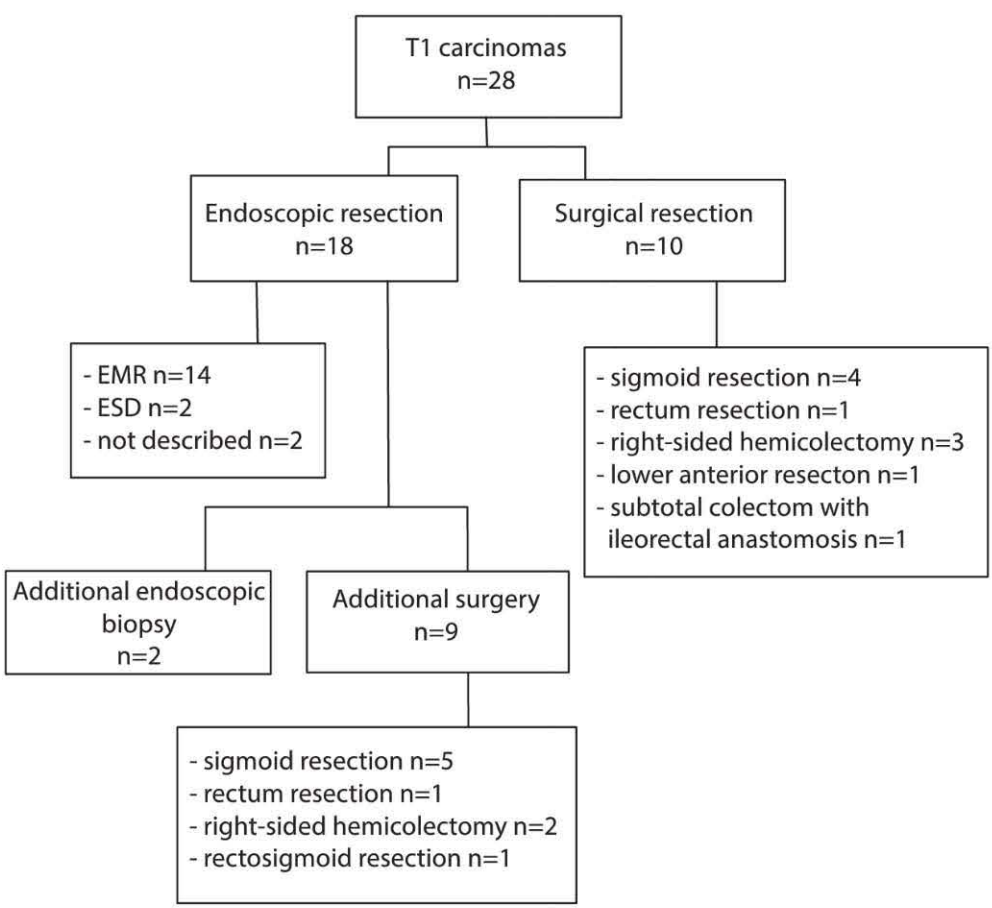


Figure 1. Resection techniques used in pT1 colorectal carcinomas.

Tumour morphology was significantly different between T1 CRCs initially treated by endoscopy and those that were directly referred for a surgical operation, p=0.02, Table 1. In T1 CRCs initially resected by endoscopy, polyp morphology was pedunculated in 56%, flat in 6%, sessile in 3%, and polyp morphology was not reported in 22%. In those referred for surgical resection, none were pedunculated or flat, 50% was sessile, and in 50% polyp morphology was not reported. Of the eighteen pT1 CRC for which an endoscopic resection was performed, 17% had a diameter ≥ 20 mm compared to 50% in those initially treated by surgery, p=0.06. The choice between an endoscopic or surgical resection was not significantly different for males or patients > 60 years compared to females and patients < 60 years of age (p=0.32 and 0.45; Table 1).

In thirteen (46%) carcinomas of all patients with a pT1 CRC, a malignancy was correctly suspected by the endoscopist based on the endoscopic appearance of the lesion at initial screening colonoscopy. In the other fifteen (54%) patients the lesion was removed without any suspicion of a malignant polyp. Patients with lesions in which the malignancy was correctly identified as a T1 CRC at colonoscopy underwent significantly more surgical operations as initial treatment (69%) than those in which the lesion was not identified as being malignant (7%), p 0.001.

Endoscopic resection techniques of pT1 colorectal carcinomas

Endoscopic resections were performed by using various resection techniques including EMR and ESD, Figure 1. The endoscopic en bloc resection rate was 94% (17/18). With regard to the endoscopic complication rate, one patient suffered from a colonic perforation subsequent to a primary endoscopic resection (ESD), for which a right-sided hemicolectomy was performed. Of all patients with a pT1 CRC that were directly referred for surgery, endoscopic biopsy was performed at screening colonoscopy to provide histological proof of malignancy in nine out of ten (90%). The other patient was diagnosed with FAP at screening colonoscopy and underwent subsequent surgery. The resected colon in this patient showed pT1 CRC.

Histopathologic aspects of pT1 colorectal carcinomas

Of all 28 T1 CRCs, three (11%) had unfavourable histopathologic characteristics: one (4%) was poorly differentiated and two (7%) showed lymphovascular invasion (Table 2). Endoscopic R0 resection with resection margin >1mm free of cancer was achieved in six (33%) of eighteen pT1 CRCs resected by endoscopy. Of all T1 CRC resected by endoscopy, histopathologic criteria for a curable endoscopic resection (R0 resection and favourable histopathologic features) were met in 28% (5/18).

Table 2. *Histopathologic characteristics of T1 carcinomas.*

	T1 carcinomas
	(n = 28)
Differentiation grade (n, %)	
Well differentiated	2 (7)
Moderately differentiated	17 (61)
Poorly differentiated	1 (4)
Not described	8 (29)
Lymphovascular invasion present (n, %)	2 (7)
Endoscopic resection margins*	
Free >1mm	6 (33)
Free ≤1mm	4 (15)
Not free or could not be assessed	8 (59)

*Surgical resection not taken into account (n=18)

Additional surgery

Of the eighteen patients with a T1 carcinoma that were initially treated with endoscopic resection, additional surgery was performed in nine (50%) patients, Figure 1. Reasons for additional surgery in these patients were R1 resection or resection margin < 1 mm (n=3), histopathologic feature missing (n=3), lymphovascular invasion (n=1), and colon perforation (n=1). In one patient the reason for additional surgery was unclear.

Discussion

In our FIT-based CRC screening cohort, one third of all patients with a screen-detected colorectal carcinoma had a pT1 CRC. Of these patients, 68% had eventually undergone surgery, either as initial treatment or as additional therapy to endoscopic resection. Regarding the endoscopic aspects of the pT1 CRCs, we found that all cancers with a pedunculated polyp morphology were initially resected by endoscopy, whereas sessile pT1 CRCs were significantly more treated by surgery. Also, there was a trend that larger polyps were treated more often by surgery than endoscopic resection. Unfavourable histopathologic features (lymphovascular invasion or poor differentiation) were found in only 11% of all pT1 CRCs, indicating that a substantial part of patients with these lesions could have been prevented from surgery according to the international guidelines.

In accordance with these guidelines, the initial treatment of choice in practice was endoscopic resection in the majority of all patients with a pT1 CRC. Interestingly, surgery was significantly more often chosen as initial treatment in patients in which the lesion was correctly identified as malignant by the screening endoscopist than in patients in which the lesion was not suspected to

be malignant. Besides histopathological features, other important factors that determine whether a lesion can be adequately resected by endoscopy include tumour size, location, and spreading of the tumour, which have been associated with difficult endoscopic resections[23]. In line with literature, our results showed that all T1 CRCs with a pedunculated tumour morphology were initially treated by endoscopic resection whereas the majority of sessile malignancies were referred for surgical resection. This finding suggests that tumour morphology is an important factor in the decision making of endoscopists to perform an endoscopic resection or to refer to surgery. Also, there was a trend that larger and more proximal located T1 carcinomas were treated more often by primary surgery than endoscopic resection.

In this study, endoscopic resections were performed by EMR in the majority of T1 CRCs. Preferable, resections for early-stage neoplasia are performed en bloc as en bloc resections allow full and accurate histological evaluation by pathologists and recurrence is less likely than for piecemeal resections[24-27]. Colorectal lesions smaller than 20 mm in diameter and pedunculated polyps can be treated effectively and safely by EMR[28]. ESD is recommended for larger-, flat- and sessile lesions[11,27]. Although, we showed that en bloc resection rates were high, additional surgery was required in a substantial group of patients that were initially treated by endoscopic resection often due to incomplete resection margins. This finding underlines the importance of adequate use of endoscopic resection techniques. In case limited endoscopic resection techniques are available at screening site, referral to specialized hospitals can be considered for early neoplasia's that require more advanced resection techniques such as ESD.

In order to accurately remove early CRC by endoscopy, it is mandatory that endoscopists distinguish malignant lesions from benign lesions. Several polyp classification systems have been published to identify high-risk polyps by predicting depth of invasion and risk of nodal involvement, such as the Kudo classification, the NICE classification and Paris classification[22,29,30]. In this study, only half of the T1 CRCs were identified as lesions with malignant endoscopic features. This is in line with a recent publication, which showed that sensitivity for the diagnosis of T1 CRCs based on endoscopic images is poor[6]. Our finding indicates that effort should be made to increase the awareness of endoscopists to identify malignant polyps from benign polyps and carefully consider subsequent resection techniques.

To the best of our knowledge, prevalence and current treatment of T1 CRCs that can be adequately treated by endoscopy have not previously been evaluated in a screening setting. Strong points

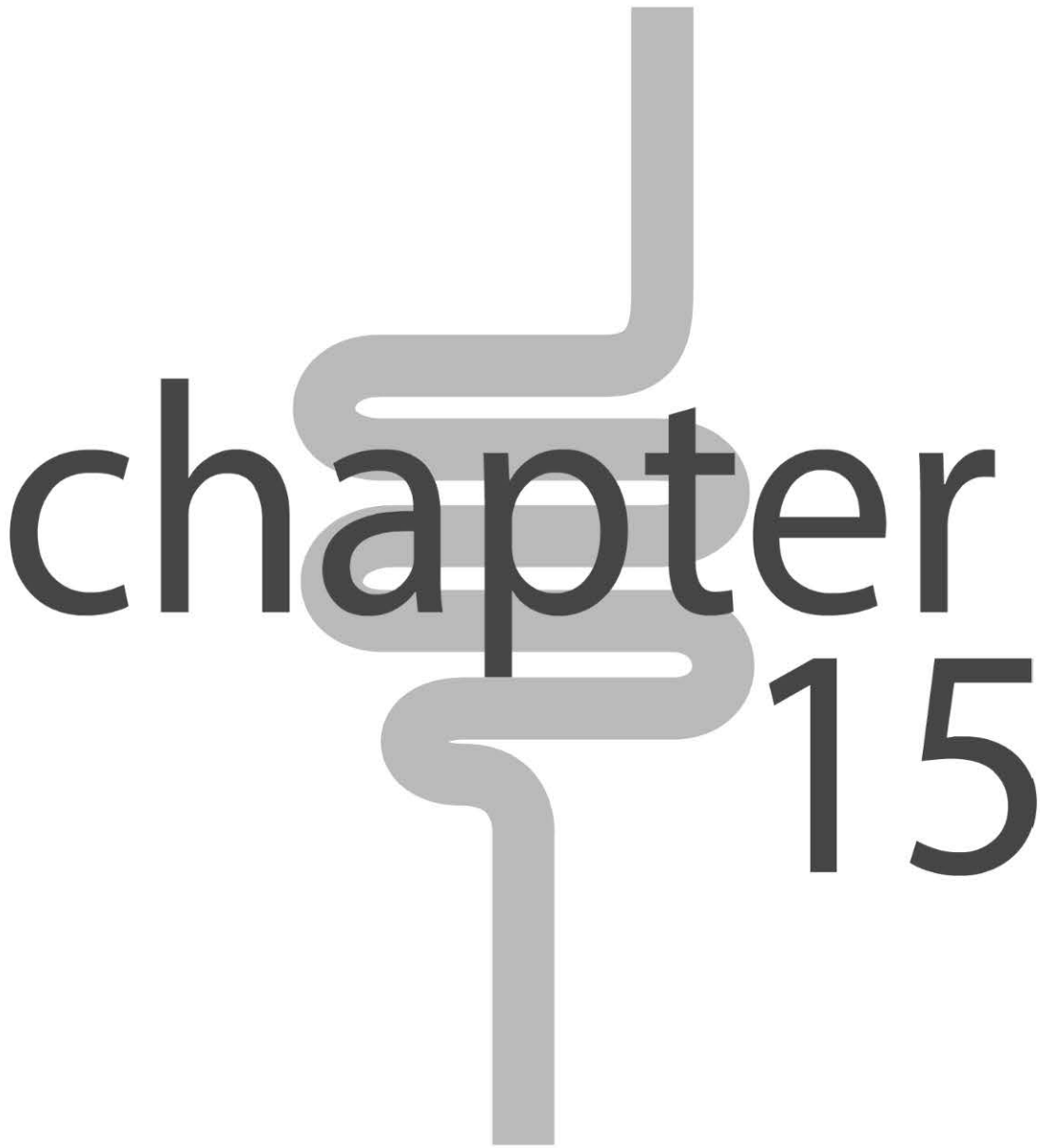
of our study are that the initial study was prospectively designed and that data were available from multiple hospitals (secondary, teaching and tertiary), resulting in a reliable reflection of a nationwide screening program. The fact that some of these hospitals may have had poor access to advanced endoscopic resection techniques, may have influenced the choice of treatment and resection techniques used. Thanks to these strengths, our findings provide insight into the current practice of identification and endoscopic removal of T1 CRCs in an average screening setting in the Netherlands. Our study has some limitations. The main limitation is the small number of T1 CRCs in this cohort and future research is needed to confirm our findings in larger screening populations. Nevertheless, our study provides a first impression on this subject. Another limitation is the wide time frame in which our CRC screening program was implemented. Since endoscopists awareness of the identification of malignant polyps and of choosing adequate resection techniques might have improved over the years, this could have affected our results.

In conclusion, a substantial proportion of pT1 CRCs detected in a FIT-based CRC screening is amiable for endoscopic resection according to the guidelines. In practice, the majority of pT1 CRCs is treated by surgical resection, either as initial treatment or subsequent to an endoscopic resection. Adequate endoscopic resection as approach to invasive T1 CRC must balance surgical morbidity against the potential benefits of endoscopic resection. Effort should be made to improve awareness of screening endoscopists to identify T1 CRCs and make informed decisions on the resection method used.

References

1. Schreuders EH, Ruco A, Rabeneck L et al. Colorectal cancer screening: a global overview of existing programmes. Gut 2015, DOI: gutjnl-2014-309086 [pii] 10.1136/gutjnl-2014-309086:
2. Logan RF, Patnick J, Nickerson C et al. Outcomes of the Bowel Cancer Screening Programme (BCSP) in England after the first 1 million tests. Gut 2012; 61: 1439-1446
3. Paimela H, Malila N, Palva T et al. Early detection of colorectal cancer with faecal occult blood test screening. Br J Surg 2010; 97: 1567-1571
4. Kubisch CH, Crispin A, Mansmann U et al. Screening for Colorectal Cancer is Associated with Lower Disease Stage: A Population-Based Study. Clin Gastroenterol Hepatol 2016, DOI: S1542-3565(16)30068-4 [pii]
5. 10.1016/j.cgh.2016.04.008:
6. Moreno CC, Mittal PK, Sullivan PS et al. Colorectal Cancer Initial Diagnosis: Screening Colonoscopy, Diagnostic Colonoscopy, or Emergent Surgery, and Tumor Stage and Size at Initial Presentation. Clin Colorectal Cancer 2016; 15: 67-73
7. van der Vlugt M, van Doorn SC, Wang J et al. Optical diagnosis of malignant colorectal polyps: is it feasible? Endosc Int Open 2016; 04: E778-E783
8. Wang S, Gao S, Yang W et al. Endoscopic submucosal dissection versus local excision for early rectal cancer: a systematic review and meta-analysis. Tech Coloproctol 2016; 20: 1-9
9. Beaton C, Twine CP, Williams GL et al. Systematic review and meta-analysis of histopathological factors influencing the risk of lymph node metastasis in early colorectal cancer. Colorectal Dis 2013; 15: 788-797
10. Choi DH, Sohn DK, Chang HJ et al. Indications for subsequent surgery after endoscopic resection of submucosally invasive colorectal carcinomas: a prospective cohort study. Dis Colon Rectum 2009; 52: 438-445
11. Rabeneck L, Paszat LF, Hilsden RJ et al. Bleeding and perforation after outpatient colonoscopy and their risk factors in usual clinical practice. Gastroenterology 2008; 135: 1899-1906, 1906 e1891
12. Pimentel-Nunes P, Dinis-Ribeiro M, Ponchon T et al. Endoscopic submucosal dissection: European Society of Gastrointestinal Endoscopy (ESGE) Guideline. Endoscopy 2015; 47: 829-854
13. Tanaka S, Kashida H, Saito Y et al. JGES guidelines for colorectal endoscopic submucosal dissection/endoscopic mucosal resection. Dig Endosc 2015; 27: 417-434
14. Arezzo A, Passera R, Marchese N et al. Systematic review and meta-analysis of endoscopic submucosal dissection vs endoscopic mucosal resection for colorectal lesions. United European Gastroenterol J 2016; 4: 18-29
15. Chao G, Zhang S, Si J. Comparing endoscopic mucosal resection with endoscopic submucosal dissection: the different endoscopic techniques for colorectal tumors. J Surg Res 2016; 202: 204-215
16. Kapidzic A, Grobbee EJ, Hol L et al. Attendance and yield over three rounds of population-based fecal immunochemical test screening. Am J Gastroenterol 2014; 109: 1257-1264
17. Kapidzic A, van Roon AH, van Leerdam ME et al. Attendance and diagnostic yield of repeated two-sample faecal immunochemical test screening for colorectal cancer. Gut 2015, DOI: gutjnl-2014-308957 [pii]10.1136/gutjnl-2014-308957:
18. Marijnen CAM, van Vegchel T, Steutel A et al. Richtlijn colorectaal carcinoom 3.0 http://www.mdl.nl/uploads/240/1479/Colorectaalcarcinoom_def.pdf.nl
19. Davila RE, Rajan E, Adler D et al. ASGE guideline: the role of endoscopy in the diagnosis, staging, and management of colorectal cancer. Gastrointest Endosc 2005; 61: 1-7

20. Federici A, Giorgi Rossi P, Borgia P et al. The immunochemical faecal occult blood test leads to higher compliance than the guaiac for colorectal cancer screening programmes: a cluster randomized controlled trial. J Med Screen 2005; 12: 83-88
21. Schlemper RJ, Riddell RH, Kato Y et al. The Vienna classification of gastrointestinal epithelial neoplasia. Gut 2000; 47: 251-255
22. Bujanda L, Cosme A, Gil I et al. Malignant colorectal polyps. World J Gastroenterol 2010; 16: 3103-3111
23. The Paris Workshop. The Paris endoscopic classification of superficial neoplastic lesions: esophagus, stomach, and colon: November 30 to December 1, 2002. Gastrointest Endosc 2003; 58: S3-43
24. Hori K, Uraoka T, Harada K et al. Predictive factors for technically difficult endoscopic submucosal dissection in the colorectum. Endoscopy 2014, DOI: 10.1055/s-0034-1377205:
25. Church JM. Avoiding surgery in patients with colorectal polyps. Dis Colon Rectum 2003; 46: 1513-1516
26. Puli SR, Kakugawa Y, Saito Y et al. Successful complete cure en-bloc resection of large nonpedunculated colonic polyps by endoscopic submucosal dissection: a meta-analysis and systematic review. Ann Surg Oncol 2009; 16: 2147-2151
27. Oka S, Tanaka S, Saito Y et al. Local recurrence after endoscopic resection for large colorectal neoplasia: a multicenter prospective study in Japan. Am J Gastroenterol 2015; 110: 697-707
28. Repici A, Hassan C, De Paula Pessoa D et al. Efficacy and safety of endoscopic submucosal dissection for colorectal neoplasia: a systematic review. Endoscopy 2012; 44: 137-150
29. Repici A, Pellicano R, Strangio G et al. Endoscopic mucosal resection for early colorectal neoplasia: pathologic basis, procedures, and outcomes. Dis Colon Rectum 2009; 52: 1502-1515
30. Kudo S, Hirota S, Nakajima T et al. Colorectal tumours and pit pattern. J Clin Pathol 1994; 47: 880-885
31. Hewett DG, Kaltenbach T, Sano Y et al. Validation of a simple classification system for endoscopic diagnosis of small colorectal polyps using narrow-band imaging. Gastroenterology 2012; 143: 599-607 e591


chapter
15

Comparison of cecal intubation and adenoma detection between hospitals can provide incentives to improve quality of colonoscopy

Esmée J. Grobbee, Tim D.G. Belderbos*, Martijn G.H. van Oijen, Maarten A.C. Meijssen, Rob J.T.H. Ouwendijk, Thjon J. Tang, Frank ter Borg, Peter van der Schaar, Doris M. Le Fèvre, Merijn T. Stouten, Onno van der Galiën, Theo J. Hiemstra, Wouter H. de Vos tot Nederveen Cappel, Pieter C.J. ter Borg, Manon C.W. Spaander, Leon M.G. Moons, Ernst J. Kuipers, Peter D. Siersem *both authors contributed equally*

Endoscopy, 2015

Abstract

Background Cecal intubation rate (CIR) and adenoma detection rate (ADR) have been found to be inversely associated with the occurrence of post-colonoscopy colorectal cancer. Depicting differences in CIR and ADR between hospitals could provide incentives for quality improvement. We aimed to compare quality parameters of routine colonoscopies between seven hospitals in the Netherlands to determine to what extent possible differences are attributable to procedural and institutional factors.

Methods We prospectively included consecutive patients undergoing colonoscopy between November 2012 and January 2013 in two academic and five non-academic medical hospitals. Patients with inflammatory bowel disease or hereditary colorectal cancer syndromes were excluded. Main outcome measures were CIR and ADR.

Results A total of 3,129 patients were included (46% male; mean age 59±15 years). The majority of patients (86%) had a Boston Bowel Preparation Scale (BBPS) score ≥6. Mean CIR was 95%, ranging from 89% to 99% between hospitals ($p<0.001$). After adjustment for casemix (age, gender, American Society of Anesthesiologists-score and indication for colonoscopy), factors associated with CIR were hospital and a BBPS ≥6. Mean ADR was 32% and varied between hospitals, ranging from 25-47% ($p<0.001$). Independent predictors for ADR were hospital, a BBPS ≥6 and cecal intubation. By combining CIR and ADR per hospital we developed a colonoscopy quality indicator (CQI) that can be used by hospitals to stimulate quality improvement.

Conclusion Differences in quality of colonoscopy between hospitals can be demonstrated using CIR and ADR. As both indicators are affected by institution and bowel preparation a comparison between hospitals based on the newly developed CQI could assist in further improving the quality of colonoscopy

Introduction

Screening strategies employing colonoscopy for detection and removal of precursors of colorectal cancer (CRC) effectively reduce CRC-related mortality[1,2]. Apart from participation grades, the efficacy of population-based CRC screening depends on the quality of colonoscopy.

Colonoscopy quality is ultimately reflected by a reduction in the incidence of CRC following colonoscopy. However, measurement of post-colonoscopy CRC (PC-CRC) is cumbersome and does not allow direct feedback. Several procedural indicators have been suggested for monitoring quality[3]. Two recent studies have shown that the adenoma detection rate (ADR) of endoscopists was inversely associated with the risk of PC-CRC [4,5] and PC-CRC related death.[5] In another study, patients dying from CRC had a lower probability of having undergone a previous complete colonoscopy than matched controls[6]. In addition, a substantial proportion of colorectal tumors originates from the right sided colon,which underlines the importance of performing a complete colonoscopy, as measured by cecal intubation rate (CIR)[7].

ADR and CIR have been reported to vary between hospitals depending on casemix and institutional or procedural factors[8-10]. Casemix is determined by non-modifiable patient characteristics, such as age, gender, comorbidity and indication for colonoscopy. In order to provide a useful incentive for hospitals to improve quality of colonoscopy, a comparison of CIR and ADR between institutions should be able to detect differences that are independently affected by modifiable factors. In this study we compared the quality of routine colonoscopy between seven hospitals in the Netherlands to determine to what extent detected differences in CIR and ADR were indeed attributable to procedural and institutional factors.

Methods

Registration of colonoscopy data

All colonoscopies performed between November 1st, 2012 and January 10th, 2013 in two academic medical centers and five large non-academic medical hospitals were prospectively registered. Patient characteristics, i.e. age, gender and American Society of Anesthesiologists (ASA) score were obtained from the electronic medical records. All endoscopists filled out data on the indication for colonoscopy, the type and dose of sedation used, quality of bowel preparation, cecal intubation, detection and removal of polyps, results from pathology reports and complications.

Definitions

Endoscopists were categorized as gastroenterologists, fellows-in-training for gastroenterology or nurse endoscopists. Indications for colonoscopy were grouped into five categories: (1) anemia or abdominal symptoms; (2) overt or occult rectal blood loss; (3) screening or a positive family history for CRC; (4) surveillance after CRC or colorectal adenoma(s) and (5) other. The latter category consisted mostly of patients with liver metastases or other abnormalities found during imaging. Patients with inflammatory bowel disease or hereditary CRC or polyposis syndromes were excluded. Bowel preparation was scored according to the Boston Bowel Preparations Scale (BBPS). Adequate bowel preparation was defined as a BBPS ≥6.[11] Sedation was provided using midazolam, propofol and/or opioid analgesics. Cecal landmarks were noted in the colonoscopy report; photographic documentation was not routinely obtained. The unadjusted CIR was defined as the proportion of colonoscopies in which the cecum was visualized, irrespective of reasons for not intubating the cecum. The adjusted CIR was calculated by excluding colonoscopies in which the endoscopist made the decision not to intubate the cecum because of severe colitis, colonic obstruction or therapeutic targets not necessitating cecal intubation.[12] Colonoscopies with poor bowel preparation were included in the adjusted cecal intubation rate, as preparation is considered to be part of the colonoscopy practice in hospitals. Complications were subdivided into bleeding (only taken into account if not stopped spontaneously or by an intervention during the colonoscopy), perforation and post-polypectomy syndrome.

Factors comprised in the term casemix are patients' age, gender, ASA score and indication for colonoscopy. Correction for casemix was performed by taking these factors into account when analyzing the association of modifiable factors with the CIR and ADR. Modifiable factors were institution, i.e. the hospital where the colonoscopy was performed, and procedural factors, consisting of the type of endoscopist (gastroenterologists, fellows, nurse endoscopists), use of conscious sedation and BBPS score.

Primary outcomes were unadjusted CIR, adjusted CIR and ADR, defined as the proportion of procedures in which one or more adenomas were found. Secondary outcomes were BBPS score, mean number of adenomas per procedure (MAP), mean number of adenomas per positive procedure (MAP+), and complications.

We constructed a colonoscopy quality indicator (CQI) by plotting the adjusted CIR and the ADR per hospital. The sizes of the dots represent the number of colonoscopies performed in each hospital.

The position of each dot can be compared to other hospitals and to predetermined thresholds. For the adjusted CIR a minimum of 95% is generally accepted3, but there is no unequivocal minimum for ADR during routine colonoscopies. Therefore, we chose to draw a line representing approximately the average ADR, creating a visual target of performing better than average.

Statistical analysis

Data are presented as percentages for categorical variables and means (including standard deviation) or medians (including ranges) for continuous variables, according to the nature of their distribution. Differences between groups were tested using the chi-squared test for categorical variables and the t-test for normally distributed continuous variables. We performed a logistic regression analysis to identify factors associated with adjusted CIR and ADR. Factors with a p-value <0.10 in univariable analysis, were included in a multivariable model. A two-sided p-value of <0.05 was considered statistically significant. Analyses were performed using SPSS 20.0 statistics software (IBM, Chicago, Ill.).

Informed consent

This study was exempted from patients' informed consent as determined by the Medical Ethical Committee of the UMC Utrecht in accordance with the Medical Research Involving Human Subjects Act.

Results

A total of 3,129 patients underwent colonoscopy during the study period. Mean age was 59 ± 15 years and 46% were male (Table 1). In the majority of cases (63%) the indication for colonoscopy was anemia or abdominal symptoms. Ninety-one percent of patients had an ASA-score of 1 or 2. Conscious sedation was used in more than 90% of cases. Split dose bowel preparation was common practice in all 7 hospitals. Gastroenterologists performed most colonoscopies (60%), followed by gastroenterology fellows (24%). The number of colonoscopies per hospital ranged between 124 and 793 (median 421).

Eighty-six percent (n=2,697) of all procedures were performed in adequately prepared colons, with a median BBPS of 9 (interquartile range 6–9) (Table 2). The unadjusted CIR was 95%. Reasons for not intubating the cecum were inadequate preparation (n=58, 36%), stenosis (n=35, 22%), technical difficulty (n=24, 15%), obstructing tumor (n=15, 9%), therapeutic goal for which cecal intubation

was not required (n=11, 7%), pain (n=9, 6%), diverticulosis (n=8, 5%) or severe inflammation (n=2, 1%). The adjusted CIR was 96%.

Table 1. *Baseline characteristics*

	Number of patients (%)
	(N= 3,129)
Age, mean ± SD (years)	59 ± 15
Male gender	1,423 (45.5)
Indication	
Anemia / abdominal symptoms[1]	1,972 (63.0)
Rectal (occult) blood loss	226 (7.2)
Family history of CRC	263 (8.4)
Surveillance after CRC / adenoma	626 (20.0)
Other[2]	42 (1.3)
ASA-score	
1	1,784 (57.0)
2	1,061 (33.9)
3	160 (5.1)
4	6 (0.2)
Unknown	118 (3.8)
Conscious sedation	
Yes	2,840 (90.8)
No	158 (5.0)
Unknown	131 (4.2)
Endoscopist	
Gastroenterologist	1,870 (59.8)
Gastroenterology fellow	747 (23.9)
Nurse endoscopist	512 (16.4)
Organization	
1	339 (10.8)
2	245 (7.8)
3	124 (4.0)
4	639 (20.4)
5	568 (18.2)
6	421 (13.5)
7	793 (25.3)

1. Includes changes in bowel habit.
2. Includes liver metastases, abnormalities found with other imaging modalities or diverticular disease

One or more polyps were detected in 45% of all colonoscopies. The ADR was 32% with a mean number of 0.6 (± 1.2) adenomas per procedure for all procedures combined (MAP) and 1.9 (±1.5) for procedures in which at least one adenoma was found (MAP+). Complications were observed in 0.6% (n=19) of procedures, of which 47% (n=9) were bleedings.

A significant variability for CIR and ADR was found between hospitals (Table 3). Adequate bowel preparation varied between 79% and 98% (p<0.001). Unadjusted CIR ranged from 89% to 99% (p<0.001). Adjusted CIR was above 90% in all participating hospitals and was also significantly different between hospitals (p<0.001). ADR varied from 25% to 47% (P<0.001). CIR and ADR were combined in a colonoscopy quality indicator (CQI) to depict differences between hospitals, taking into account the number of performed procedures per hospital (Figure 1).

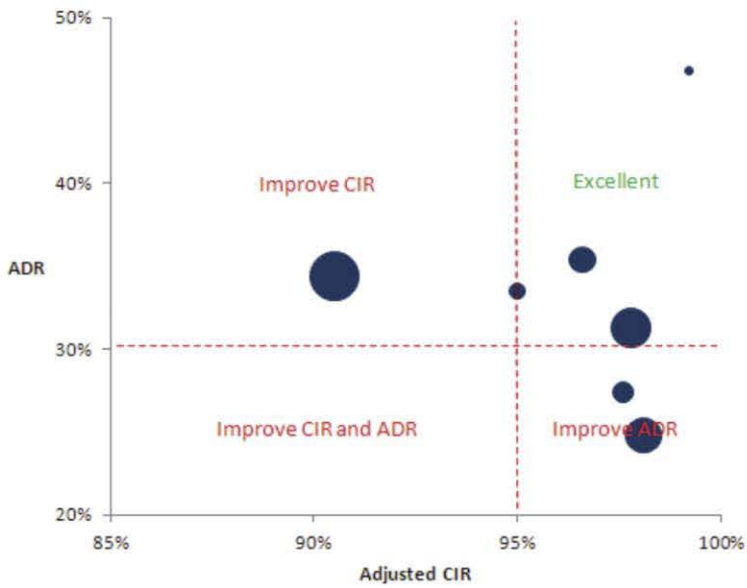


Figure 1. *Quality indicators (adenoma detection rate (ADR) and cecal intubation rate (CIR)) per hospital.*

Complication rates were non-significantly different between hospitals (p=0.074). ADR or CIR were not higher in academic centers or centers with a higher volume. A comparison between the different types of endoscopists did not show differences in unadjusted or adjusted CIR (p=0.447 and p=0.501, respectively)(Table 3). Without correction for casemix, both nurse endoscopists (36%) and fellows (34%) had a higher ADR than gastroenterologists (30%, p=0.008). No association was found between the number of procedures per endoscopist and CIR or ADR. Complications occurred more frequently after colonoscopies performed by fellows (p=0.042).

In univariable analysis, hospital, age, ASA-score, indication for colonoscopy and BBPS were identified as factors associated with adjusted CIR. Hospital, ASA-score, indication and BBPS remained significantly associated with the adjusted CIR in multivariable analysis (Table 4). In patients with an ASA score >1, the cecum was less frequently intubated. If patients underwent colonoscopy for surveillance, the procedure was more often completed than in case of abdominal complaints or anemia. CIR was significantly higher in patients with adequate bowel preparation.

Table 2. *Quality indicators for all hospitals.*

Indicators	Results (%)
	(N=3,129)
BBPS ≥6	
Yes	2,697 (86.2)
No	310 (9.9)
Unknown	122 (3.9)
BBPS, median (interquartile range)	9 (6-9)
CIR	
Yes	2,967 (94.8)
No	162 (5.2)
Adjusted CIR	
Yes	2,967 (95.7)
No	134 (4.3)
PDR	1,413 (45.2)
ADR	996 (31.8)
MAP	0.60 ± 1.22
MAP+	1.89 ± 1.48
Complications	19 (0.6)
Bleeding	9 (0.3)
Post-polypectomy syndrome	2 (0.1)
Perforation	1 (<0.1)
Other/unknown	7 (0.2)

Hospital, type of endoscopist, gender, age, ASA-score, indication for colonoscopy, BBPS and unadjusted CIR were all associated with ADR in univariable analysis. In multivariable analysis, an association was shown for hospital, gender, age, indication, BBPS and unadjusted CIR (Table 5). Male gender and older age were associated with a higher ADR. One or more adenomas were more frequently found if the indication for colonoscopy was a positive family history, rectal blood loss or surveillance after previous adenoma(s) or CRC, in comparison to anemia or abdominal symptoms. Both adequate bowel preparation and cecal intubation increased the likelihood of detecting adenomas. After adjustment for casemix, there were no significant differences in ADR between gastroenterologists, fellows and nurse endoscopists.

Table 3. *Outcomes per hospital and type of endoscopist*

Hospital	Number, n	BBPS ≥6, %	CIR, %	Adj. CIR, %	ADR, %	Complications, %
1	339	79.0	96.2	97.6	27.4	1.2
2	245	87.0	93.1	95.0	33.5	1.6
3	124	88.7	99.2	99.2	46.8	0.0
4	639	97.3	97.5	97.8	31.3	0.8
5	568	95.8	98.1	98.1	24.8	0.2
6	421	97.6	95.2	96.6	35.4	0.0
7	793	80.2	89.4	90.5	34.4	0.6
		p<0.001	*p<0.001*	*p<0.001*	*p<0.001*	*p=0.074*
Gastroenterologist	1870	90.1	95.3	95.5	29.8	0.5
Fellow	747	85.7	94.7	95.7	34.0	1.2
Nurseendoscopist	512	93.9	96.2	96.5	36.1	0.2
		p<0.001	*p=0.447*	*p=0.629*	*p=0.008*	*p=0.042*

Discussion

In this prospective cohort of 3,129 routine colonoscopies in seven hospitals we demonstrated an overall CIR of 95% and an ADR of 32%, respectively. Our results indicate a significant variability in both CIR and ADR between different hospitals. Both CIR and ADR were affected by hospital and bowel preparation, independent of casemix variation. We developed a colonoscopy quality indicator (CQI), which shows both CIR and ADR in a matrix that can be used for quality assessment and may assist in improving colonoscopy performance of hospitals and individual endoscopists.

CIR and ADR in this study were higher than in previous studies, but the variability between hospitals was comparable. Mean CIR and ADR as reported in several studies including routine colonoscopies have been found to vary from 83 to 91% and from 18 to 26%, respectively[8, 13-18]. De Jonge et al. reported an unadjusted CIR varying from 81 to 96% and an ADR varying from 13 to 32% in twelve Dutch academic and non-academic hospitals whereas Harris et al. found a CIR of 69 to 98% and an ADR of 8 to 27% in 21 experienced centers across Europe and Canada[15-18]. The variation ADR was even higher in the study by Radaelli et al, ranging from 6 to 46% in 116 Italian endoscopy centers participating in a nationwide registration study for routine colonoscopies[17]. Different studies among screening populations have shown slightly higher CIRs and ADRs, but a comparable variability[10,19-21]. The slightly better results in the current study may at least partly be explained by a specific focus on colonoscopy practice in the centers participating in this quality initiative.

***Table 4**. Factors associated with adjusted cecal intubation rate*

	Unadjusted OR (95% CI)	Adjusted OR (95% CI)
Hospital		
1	4.25 (2.03-8.93)	7.74 (2.68-22.31)
2	1.98 (1.06-3.72)	1.17 (0.57-2.40)
3	12.84 (1.77-93.21)	14.66 (1.78-120.66)
4	4.65 (2.60-8.31)	1.75 (0.89-3.43)
5	5.29 (2.78-10.05)	2.90 (1.33-6.35)
6	2.99 (1.67-5.36)	1.99 (0.86-4.61)
7	Reference	Reference
Type of endoscopist		
Gastroenterologist	Reference	
Fellow	1.04 (0.69-1.58)	
Nurse-endoscopist	1.29 (0.77-2.17)	
Male gender	1.34 (0.94-1.91)	
Age, per 10 years increase	0.78 (0.69-0.89)	0.92 (0.79-1.07)
ASA-score		
1	Reference	
2	0.45 (0.31-0.65)	
3	0.25 (0.14-0.46)	
4	0.09 (0.01-0.85)	
Indication		
Anemia / abdominal symptoms	Reference	
Rectal (occult) blood loss	1.70 (0.78-3.71)	
Family history of CRC	4.70 (1.48-14.93)	
Surveillance	1.85 (1.11-3.08	
Sedation	0.74 (0.27-2.04)	
Adequate BBPS	15.61 (10.56-23.08)	14.64 (9.25-23.16)

Table 5. *Factors associated with adenoma detection rate.*

	Unadjusted OR (95% CI)	**Adjusted OR (95% CI)**
Hospital		
1	0.72 (0.54-0.95)	0.83 (0.58-1.19)
2	0.96 (0.71-1.30)	0.95 (0.67-1.35)
3	1.67 (1.14-2.45)	1.31 (0.83-2.07)
4	0.87 (0.70-1.08)	0.73 (0.56-0.93)
5	0.63 (0.50-0.80)	0.55 (0.41-0.74)
6	1.04 (0.81-1.34)	0.77 (0.56-1.04)
7	Reference	Reference
Type of endoscopist		
Gastroenterologist	Reference	Reference
Fellow	1.21 (1.01-1.46)	1.12 (0.89-1.42)
Nurse-endoscopist	1.33 (1.09-1.64)	1.27 (0.98-1.63)
Male gender	1.78 (1.53-2.07)	1.73 (1.47-2.05)
Age, per 10 years increase	1.51 (1.43-1.61)	1.48 (1.38-1.59)
ASA-score		
1	Reference	Reference
2	1.75 (1.49-2.06)	1.13 (0.93-1.37)
3	2.01 (1.44-2.79)	1.37 (0.94-1.99)
4	1.36 (0.25-7.44)	1.86 (0.28-12.47)
Indication		
Anemia / abdominal symptoms	Reference	Reference
Rectal (occult) blood loss	1.82 (1.37-2.41)	1.63 (1.16-2.29)
Family history of CRC	1.14 (0.86-1.51)	1.42 (1.02-1.97)
Surveillance	2.34 (1.98-2.88)	1.70 (1.38-2.09)
Sedation	1.26 (0.88-1.80)	
Adequate BBPS	1.56 (1.19-2.05)	1.84 (1.33-2.55)
Cecal intubation	1.81 (1.23-2.66)	1.99 (1.24-3.20)

The adjusted CIR in this study was found to be associated with ASA score and indication for colonoscopy as well as bowel preparation and hospital, which are both modifiable factors. Previous studies concluded that not only adequate bowel preparation is associated with CIR, but also the use of sedation [9,12,22,23]. One other study found no association between sedation and CIR and only a non-significant (p=0.07) association between bowel preparation and CIR[15]. In our study, the use of sedatives was not related to adjusted CIR. The general Dutch practice is to perform colonoscopy under conscious sedation, unless the patient prefers not to do so. This may have resulted in confounding by indication.

As expected, our study showed that more non-modifiable factors were associated with ADR than with CIR. Age, sex and indication for colonoscopy are known to affect the ADR and adjustment for these factors is required when measuring quality[14,16,24,25]. Hospital, bowel preparation and cecal intubation were modifiable factors associated with ADR in multivariable analysis. Most previous cohort studies have reported an association between adequate bowel preparation and higher ADR[9,10,15,20,26]. It seems logical to assume that a higher CIR increases ADR, but not all previous studies have confirmed this [10,15,20]. Sedation did not affect ADR and previous studies have shown only limited and largely conflicting evidence on this subject[10,14,22,23].

We developed the CQI, which is a tool that can provide incentives for hospitals to improve quality of colonoscopy. It is important to note that the CQI does not reflect absolute quality, as the results are not corrected for casemix. Therefore, it is not fair to consider one hospital to be better than the other, solely based on a different position in the CQI-matrix. Yet, the CQI may help hospitals or individual endoscopists to detect differences compared to other hospitals or endoscopists and look into reasons for these differences. If the variation appears to result from differences in procedural or institutional factors, as was shown for bowel preparation, hospitals or endoscopists can focus on this factor to improve quality.

In this context, it is worth mentioning the seemingly outstanding performance of hospital 3 in a limited number of colonoscopies, represented by the small dot in the upper right corner of the CQI-matrix (Figure 1). Although multivariable analyses confirmed this performance to be at least partly independent of casemix, it appears that the high rates are importantly affected by differences in baseline characteristics of included patients (Supplementary Table 1). During the study period, significantly more colonoscopies were performed in patients with a positive fecal occult blood test as compared to the other hospitals.

Although it seems that there is no linear association between CIR and ADR in the CQI, it is important to emphasize that this conclusion cannot be drawn from the CQI, due to the intersection of the x- and y-axes at high percentages and the lack of correction for casemix. If the complete CQI would have been shown, i.e. with x- and y-axes ranging from 0 to 100%, one would have seen that all dots are in the upper right corner and that an estimation of co-linearity is not possible as all hospitals performed relatively well. The CIR was independently associated with the ADR in our multivariable analysis, but this cannot be directly interpreted from the CQI. It is possible to overcome this by constructing CQIs for a specific indication and age group, for example in screening populations.

The strength of this study is that our prospective registration resulted in reliable data on colonoscopies performed in a representative population of patients visiting academic and non-academic hospitals. However, limitations of this study should be addressed as well. First, several population characteristics possibly affecting ADR and/or CIR were not or incompletely known in this study. Patient's comorbidities, medication use and smoking status were not registered. ASA score as an indirect measure for these factors was registered in all patients. The indications for colonoscopy were consistently reported, but were not highly specific. Particularly, more than 60% of all patients were categorized in a combined group of anemia, abdominal symptoms or change in bowel habits, whereas one might expect this group to be heterogeneous with regard to risk of finding colorectal neoplasms. Second, some potentially important information on procedural factors was also not available. Endoscopists were aware of the recommendation that withdrawal with mucosal inspection should take at least six minutes, but withdrawal time was not recorded in this study[3]. The number of colonoscopies performed in each hospital per year and the experience of each endoscopist were not known, but no significant differences in CIR or ADR were found between different types of endoscopists. This is remarkable, as gastroenterologists are generally more experienced than fellows.

In conclusion, we investigated the use of CIR and ADR as quality indicators for colonoscopy. The CQI is a simple combination of both indicators for comparison between hospitals. It can provide information to discuss differences, aiming to improve colonoscopy quality, as CIR and ADR can be positively affected by targeting modifiable factors, of which the most important is bowel preparation. Future studies are required to establish whether implementation of competitive feedback in such a comparative way will indeed improve quality of colonoscopy and what measures are best to be taken to improve bowel preparation.

References

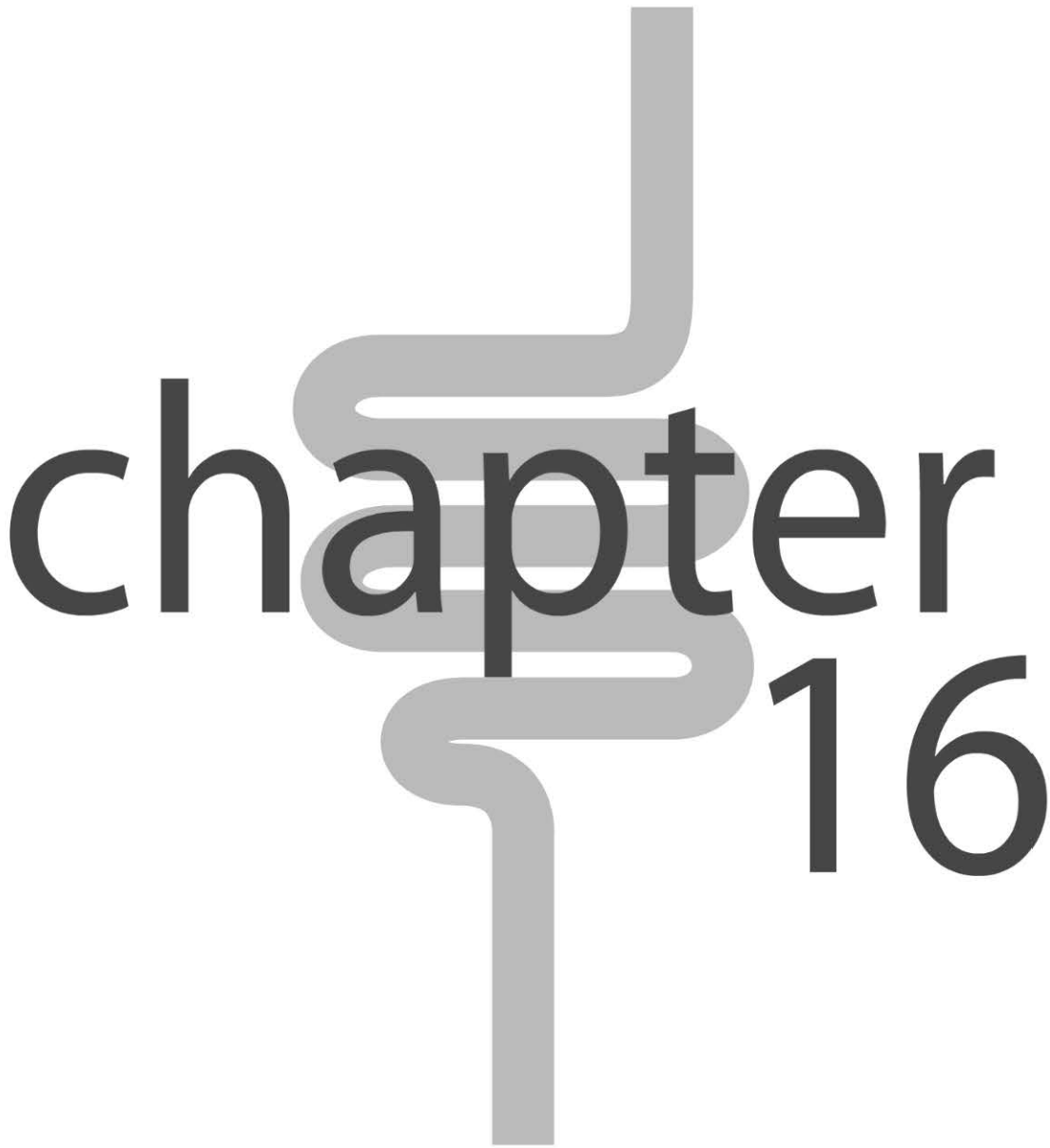
1. Zauber AG, Winawer SJ, O'Brien MJ, et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. N Engl J Med. 2012;366:687-696.
2. Amri R, Bordeianou LG, Sylla P, et al. Impact of Screening Colonoscopy on Outcomes in Colon Cancer Surgery. JAMA Surg. 2013;1-7.
3. Rex DK, Petrini JL, Baron TH, et al. Quality indicators for colonoscopy. Gastrointest Endosc. 2006;63:S16-S28.
4. Kaminski MF, Regula J, Kraszewska E, et al. Quality indicators for colonoscopy and the risk of interval cancer. N Engl J Med. 2010;362:1795-1803.
5. Corley DA, Jensen CD, Marks AR, et al. Adenoma detection rate and risk of colorectal cancer and death. N Engl J Med. 2014;370:1298-306
6. Baxter NN, Goldwasser MA, Paszat LF, et al. Association of colonoscopy and death from colorectal cancer. Ann Intern Med. 2009;150:1-8.
7. Rabeneck L, Souchek J, El-Serag HB. Survival of colorectal cancer patients hospitalized in the Veterans Affairs Health Care System. Am J Gastroenterol. 2003;98:1186-1192.
8. Ricci E, Hassan C, Petruzziello L, et al. Inter-centre variability of the adenoma detection rate: A prospective, multicentre study. Dig Liver Dis. 2013.
9. Froehlich F, Wietlisbach V, Gonvers JJ, et al. Impact of colonic cleansing on quality and diagnostic yield of colonoscopy: the European Panel of Appropriateness of Gastrointestinal Endoscopy European multicenter study. Gastrointest Endosc. 2005;61:378-384.
10. Jover R, Zapater P, Polania E, et al. Modifiable endoscopic factors that influence the adenoma detection rate in colorectal cancer screening colonoscopies. Gastrointest Endosc. 2013;77:381-389.
11. Calderwood AH, Jacobson BC. Comprehensive validation of the Boston Bowel Preparation Scale. Gastrointest Endosc. 2010;72:686-692.
12. Aslinia F, Uradomo L, Steele A, et al. Quality assessment of colonoscopic cecal intubation: an analysis of 6 years of continuous practice at a university hospital. Am J Gastroenterol. 2006;101:721-731
13. Chen SC, Rex DK. Endoscopist can be more powerful than age and male gender in predicting adenoma detection at colonoscopy. Am J Gastroenterol. 2007;102:856-861.
14. Barret M, Boustiere C, Canard JM, et al. Factors associated with adenoma detection rate and diagnosis of polyps and colorectal cancer during colonoscopy in France: results of a prospective, nationwide survey. PLoS One. 2013;8:e68947.
15. De Jonge V, Sint Nicolaas J, Cahen DL, et al. Quality evaluation of colonoscopy reporting and
16. performance in daily clinical practice. Gastrointest Endosc. 2012;75:98-106
17. Bhangu A, Bowley DM, Horner R, et al. Volume and accreditation, but not specialty, affect quality standards in colonoscopy. Br J Surg. 2012;99:1436-1444.
18. Radaelli F, Paggi S, Minoli G. Variation of quality of colonoscopy in Italy over five years: a nation-wide observational study. Dig Liver Dis. 2013;45:28-32.
19. Harris JK, Vader JP, Wietlisbach V, et al. Variations in colonoscopy practice in Europe: a multicentre descriptive study (EPAGE). Scand J Gastroenterol. 2007;42:126-134.
20. Imperiale TF, Glowinski EA, Juliar BE, et al. Variation in polyp detection rates at screening colonoscopy. Gastrointest Endosc.

2009;69:1288-1295.

21. Adler A, Wegscheider K, Lieberman D, et al. Factors determining the quality of screening colonoscopy: a prospective study on adenoma detection rates, from 12,134 examinations (Berlin colonoscopy project 3, BECOP-3). Gut. 2013;62:236-241.
22. Heitman SJ, Ronksley PE, Hilsden RJ, et al. Prevalence of adenomas and colorectal cancer in average risk individuals: a systematic review and meta-analysis. Clin Gastroenterol Hepatol. 2009;7:1272-1278.
23. Bannert C, Reinhart K, Dunkler D, et al. Sedation in screening colonoscopy: impact on quality indicators and complications. Am J Gastroenterol. 2012;107:1837-1848.
24. Radaelli F, Meucci G, Sgroi G, et al. Technical performance of colonoscopy: the key role of sedation/analgesia and other quality indicators. Am J Gastroenterol. 2008;103:1122-1130.
25. Ferlitsch M, Reinhart K, Pramhas S, et al. Sex-specific prevalence of adenomas, advanced adenomas, and colorectal cancer in individuals undergoing screening colonoscopy. JAMA. 2011;306:1352-1358.
26. Corley DA, Jensen CD, Marks AR, et al. Variation of adenoma prevalence by age, sex, race, and colon location in a large population: implications for screening and quality programs. Clin Gastroenterol Hepatol. 2013;11:172-180.
27. Lee TJ, Blanks RG, Rees CJ, et al. Longer mean colonoscopy withdrawal time is associated with increased adenoma detection: evidence from the Bowel Cancer Screening Programme in England. Endoscopy. 2013;45:20-26.
28. Barclay RL, Vicari JJ, Doughty AS, et al. Colonoscopic withdrawal times and adenoma detection during screening colonoscopy. N Engl J Med 2006;355:2533-41.
29. Moritz V, Bretthauer M, Ruud HK, et al. Withdrawal time as a quality indicator for colonoscopy - a nationwide analysis. Endoscopy 2012;44:476-81.
30. Sawhney MS, Cury MS, Neeman N, et al. Effect of institution-wide policy of colonoscopy withdrawal time> or = 7 minutes on polyp detection. Gastroenterology. 2008;135:1892-1898

PART V

general discussion and future perspectives


chapter
16

General discussion and future perspectives

Summary

This thesis aimed to explore the current status of colorectal cancer screening and the role of fecal immunochemical testing (FIT). After discussing the current status of CRC screening in part I, the second part of thesis provided a discussion on different screening methods with a focus on screening by means of FIT. Next, the use of fecal hemoglobin concentration measured by quantitative FIT, was explored for personalized screening strategies. Hereafter, this thesis dealt with quality issues in colorectal cancer screening. In this final part, the findings of our research will be summarized and future perspectives will be discussed.

Colorectal cancer screening modalities with a focus on FIT (PART II)

Various screening methods are available and can be broadly divided into two strategies; invasive, colonic imaging strategies such as colonoscopy, and non-invasive strategies such as fecal occult blood testing (FOBT)[1]. Fecal occult blood testing was originally most often done by means of guaiac FOBT (gFOBT), while at present many screening programs are changing to FIT[2]. We conducted a systematic review and meta-analysis on diagnostic test accuracy comparing FIT and gFOBT **(Chapter 2)**. The results of this review showed a higher sensitivity for both advanced neoplasia as well as colorectal cancer with FIT-screening compared to gFOBT. Specificity for the two tests were comparable.

With regard to endoscopic screening strategies, colonoscopy is considered the reference standard for diagnosing colorectal cancer (CRC). However, due to the invasiveness of the procedure, attendance rates in screening programs are generally low. Colonoscopy is recommended to be repeated every 10 years, whereas fecal occult blood testing is recommended to be repeated biennially or annually[3]. These different screening intervals make comparison between screening strategies arduous, especially taking into account that the true impact of FIT-screening is attained over multiple rounds4. Until now most studies have focused on comparing one-time endoscopic screening to only one-time FIT-screening instead of multiple rounds. For this reason, we compared colonoscopy, flexible sigmoidoscopy (FS) and FIT over four rounds of screening **(Chapter 3)**. These results showed a superior performance for FIT among invitees, yielding two-times more advanced neoplasia (AN) than screening by FS or colonoscopy, with AN being found in 4.5% (95% CI 4.2-4.9) vs. 2.3% (95% CI 2.0-2.7) and 2.2% (95% CI 1.8-2.6) respectively of all invitees. Moreover, FIT-screening had a three-times higher detection rate for CRC (0.6% vs. 0.2%; $p<0.001$). Lastly, numbers needed to scope to detect one screenee with AN were lowest for FIT with 2, followed by FS with 3 and highest for colonoscopy with 11. Nevertheless, it should be noted that the comparison of CRC detection rates between FIT and endoscopic screening is complex and results should be interpreted with caution. These results are firstly directly influenced by uptake of the screening test. Furthermore, colonoscopy allows detection and simultaneous removal of lesions in a first round, whereas FIT screening requires repeated screening at relatively short interval to provide a similar preventive impact as colonoscopy in those who participate and are willing to undergo the screening-test.

In 2014 a national CRC screening program was implemented in the Netherlands. Through a national tender it was decided to start screening with the FOB-Gold (Sentinel, Italy). Contradictory, most

previous research in the Netherlands was done using the OC-sensor (Eiken, Japan). To investigate how these two tests compare, we performed a randomized trial in the fourth round of our CRC screening cohort **(Chapter 4)**. More non-analyzable tests were seen for the FOB-Gold, due to inappropriate use by the screenee. Furthermore, while both FITs had similar positive predictive values, the positivity rate of the OC-sensor was significantly higher (7.9% vs. 6.5%; p=0.002), leading to a higher colonoscopy demand and consequently a higher diagnostic yield, despite using the same cut-off for both FITs. Because FIT brands vary widely regarding sampling size and buffer, comparing brands has proven to be arduous. In 2011 the FIT for screening working group proposed to standardize FIT Hb concentrations into µg Hb/g feces, instead of the previously used ng Hb/ml buffer[5]. However, our results illustrated that even when using this standardized measurement, differences occur regarding positivity rate. This could be explained by the fact that FITs do not directly determine hemoglobin concentration in feces, but determine hemoglobin concentration in the kit's storage buffer. This depends on both the fecal hemoglobin concentration and the amount of fecal material put into the buffer. Although manufacturers assume that the volume of fecal material sampled is stable per device, sampling volumes can in practice vary substantially. Therefore, we proposed to compare FIT brands at the same positivity rate, rather than cut-off. Moreover, a comparison on positivity rate also directly reflects the necessary colonoscopy demand, which remains an important issue in CRC screening programs[6,7]. For FIT-screening to be successful, multiple screening rounds are required and with sufficient participation rates. We showed that over three rounds of screening positivity rates as well as positive predictive values decreased **(Chapter 5)**. The positivity rate was highest in the first round (8.4%) and decreased over the second (6.0%) and third (5.7%) screening rounds (p<0.001). Consequently, together with a lower PPV over the rounds, diagnostic yield decreased (round 1: 3.3%; round 2: 1.9%; and round 3: 1.3%; p<0.001).

Hereafter, we focused solely on the participation rates over multiple FIT rounds and showed that participation remained stable over 4 screening rounds, and even increased slightly over the rounds **(Chapter 6)**. Overall, 72% participated at least once in one of the four screening rounds, while around 28% never participated. This latter group mainly involved male subjects with a low socioeconomic status. Future research should evaluate new strategies to involve this deprived group in CRC screening, as male gender and low socioeconomic status are also considered to be risk factors for the development of CRC. At present, all subgroups receive the same screening invitation strategy, while it might be necessary that invitation strategies are different between subgroups to achieve even chances of optimal outcomes[8].

In spite of high participation rates and a relatively high sensitivity for CRC of 75-85% depending on the cut-off used, the sensitivity of FIT for detection of advanced adenomas is lower and generally ranges below 50%[9-11]. Following, there is a need to increase FIT sensitivity, without losing specificity, as the latter is crucial in a screening setting. For this reason, we explored if the gut microbiome could be of clinical use in FIT screening **(Chapter 7)**. We investigated four predefined bacteria associated with the development of CRC (Escherichia coli, Fusobacterium nucleatum, Bacteroides and Fecalibacterium prausnitzii) to evaluate if it was possible to measure bacteria in FIT by quantatitative PCR (qPCR) and to see if these microbiota remained stable over the course of days. Next, we compared the qPCR results to the most advanced finding at colonoscopy. Our data showed no additional benefit for each bacterium relative to the total amount of bacteria (i.e. 16S). However, we did find that total bacterial load was significantly higher in FITs from patients with high-grade dysplasia or CRC. While these results did not aide in identifying specific bacteria to increase FIT diagnostic accuracy, the finding that the gut bacteria can be adequately measured in FIT-samples, paves the way for further research to determine the role of microbiota in FIT. Moreover, the concept remains interesting and deserves further exploration.

Using FIT as a quantitative guide in colorectal cancer screening (PART III)

FIT can be both qualitative, leading to either a positive or negative result, or quantitative resulting in the reporting of fecal Hb per µg Hb per gram feces[2]. Despite the fact that quantitative FIT is most often used in screening settings, it is invariably used and reported as a qualitative test by using a pre-specified cut-off. In this third part of the thesis, fecal Hb (fHb) concentration as measured by quantitative FIT was evaluated to use in risk prediction models.

Firstly we investigated fHb concentrations of FIT-negative screenees (i.e. fHb concentration <10 µg Hb/g feces). We showed that among negative-FIT sceenees, fHb can be of clinical aid in predicting who is most at risk of developing advanced neoplasia **(Chapter 8)**. Participants with a fHb concentration of >8 µg Hb/g have an 8-fold higher risk of developing future AN, compared to those with a fHb concentration of 0 µg Hb/g. Moreover, consecutive FIT results provide even more insight into these risks, with two FITs of 8 µg Hb/g leading up to a 14-fold risk increase of AN. We provided practical heat plots, to give an insight in these risks based on two previous negative-FIT results. These results additionally underlined the fact that men have a two-fold higher risk of developing future AN than women. This risk difference between genders can be explained by the fact that men and women have different incidence rates of CRC and consequently have different risks of developing

AN[12,13]. Despite these well-known differences, both genders are screened invariably in the same manner[12,14]. We provided an overview of these gender differences in FIT screening by analyzing screenees that performed a FIT before undergoing a screening-colonoscopy **(Chapter 9)**. Using this data we worked out two scenario; gender-tailored strategies can either level sensitivity in men and women (i.e. lower cut-off in women) or level the amount of missed lesions (i.e. lower cut-off in men.

Next, we looked at fHb concentrations of FIT-positive screenees in predicting who is at risk of a second look colonoscopy **(Chapter 10)**. Moreover, we showed that there is a high rate of second look colonoscopies in a FIT-positive patient population and that it accounts for an additional 12% in required colonoscopy resources. Finally, we demonstrated that fHb concentrations could be used to alter screening intervals to optimize screening programs and related resources **(Chapter 11)**. Combining the results from two FIT-based screening rounds, we illustrated that by lowering the cut-off in round one, the second round could theoretically be extended, resulting in similar diagnostic yield and a lesser demand on screening program-related resources.

Quality and endoscopy in colorectal cancer screening (PART IV)

In part IV quality issues in CRC screening were discussed, including quality of patient information and quality of colonoscopy. Regarding quality of patient information, we showed that there is still much to gain in providing complete and understandable information for those contemplating on CRC screening **(Chapter 12)**. We developed a website accuracy score (WAS) to score website on quality and completeness of information on CRC screening together with validated internet quality instruments. Strikingly, there was a poor correlation between quality of websites and Google rank. These results suggest that persons will miss out on high quality websites and important information. More attention should be given to the aspects of web-based information, as the use of internet for health information is continuing to grow[15].

The main goal of a screening program is ultimately the reduction in the incidence and mortality of CRC, which is driven by the number of detected CRCs and missed lesions (i.e. interval carcinomas). In **Chapter 13** we evaluated detection and miss rates of CRC over four rounds of FIT-based screening; CRC was detected in 3.9% of FIT-positive screenees, and 0.14% of FIT-negative screenees had an interval carcinoma within the FIT-screening interval. Additionally, 0.43% of those undergoing colonoscopy after a positive FIT,had a post-colonoscopy CRC. Interestingly, screen-detected CRC and FIT interval CRC were most often detected in the distal colon, while post-colonoscopy CRC were more often located in the proximal colon. This is in line with results from a gFOBT-based screening

program, and is potentially due to lower fHb concentrations from these lesions in feces6. Moreover, screen-detected CRCs had more favorable stages than non-screen-detected CRCs. Over the course of multiple screening rounds, the proportion of early CRC increases. Unfortunately, such lesions are frequently not adequately recognized at colonoscopy in FIT-positive screenees. This is an important issue, as recognition of these lesions, and a following adequate resection could potentially prevent unnecessary surgery. We showed that in our CRC-screening cohort, over half of the endoscopically resectable CRCs were not identified as such at initial colonoscopy **(Chapter 14)**. More attention should be given to the diagnosis of early CRC and the available resection methods.

Quality of a screening program is greatly dependent on the quality of colonoscopy, which is ultimately reflected by the number of post-colonoscopy CRC. However, measuring post-colonoscopy CRC is cumbersome and does not allow direct feedback. For this reason, several quality indicators have been suggested, most often used include adenoma detection rate (ADR) and cecal intubation rate (CIR)[16,17]. There is a growing body of evidence that ADR of endoscopists is inversely associated with the risk of post-colonoscopy cancers[18,19]. For this reason, we looked at quality of colonoscopy between hospitals measured by the ADR and CIR, showing marked differences between centers **(Chapter 15)**. We developed a colonoscopy quality indicator, to assist in further improving the quality of colonoscopy. Future studies should focus on these differences among hospitals to identify factors and implement strategies that can improve quality of colonoscopy. For example, in the national screening program in the Netherlands, endoscopists need to be certified to do screening colonoscopies. This certification is preceded by a training, which in part focuses on quality parameters. Such strategies could improve and uniform quality of colonoscopy among hospitals.

Future perspectives

Research on colorectal cancer screening has evolved greatly over the past decades. While improvements have been made regarding the quality and accuracy of screening strategies, the strategies itself have not changed much and still mainly consist either of colonoscopy screening or fecal occult blood test screening. The perfect screening test (e.g. with 100% sensitivity and specificity, that is non-invasive and done by all screenees) has unfortunately not been identified yet and it is unlikely that such a test will be available in the coming years. With the currently available screening methods, there is an ongoing debate on the most effective screening strategy. In the United States health care professionals tend to prefer colonoscopy. Yet, in Europe fecal occult blood testing, most frequently using FIT, seems to be favorable[2]. This can partly be explained by the low

participation rates in Europe for colonoscopy as well as by the limited colonoscopy resources. Participation is a major issue in screening programs. A screening test can be highly sensitive, but is only effective when it gets done (and when it gets done right)[20]. Furthermore, colonoscopy screening puts a large demand on colonoscopy capacity. Though limited capacity is not an issue in the United States, limitations in health care resources is likely to become an issue like in other countries. Moreover, colonoscopy screening is costly, and it is an invasive procedure from a patient's perspective. Using colonoscopy as a first-line screening method leads to a high detection rate of non-advanced adenomas at the cost of the occurrence of complications in otherwise healthy individuals. Ultimately, the numbers needed to scope to detect one person with advanced neoplasia or CRC are substantially lower for FIT compared to endoscopic screening.

In the Netherlands, a national screening program started in January 2014 using FIT. Based on previous research, a screening interval of 2 years was chosen. At present, the roll-out of the screening program is ongoing, and all inhabitants between the ages 55 and 74 years are, or will be invited for CRC screening. With many different FIT brands being available, the FOB-Gold (Sentinel, Italy) was chosen in the Netherlands through a national tender. Choosing a brand for a national screening program is difficult, and often based on a balance between evidence based efficacy and costs. All FIT brands have different sampling devices and buffers. Though manufactures claim to measure similar concentrations of hemoglobin in feces, at the same threshold positivity rates between tests vary widely. These different sampling devices and buffers, makes comparing brands burdensome. Furthermore, screening tests are being developed and improved constantly, making results on positivity rates and detection rates in previously published studies often inapplicable to current situations. The FIT for screening WEO-working group has proposed to report FIT concentrations in µg Hb/g feces, instead of the previously used ng Hb/ml buffer[5]. This in theory allows to standardize the reporting and comparison of results across programs. Nevertheless, current research has shown that even when using this standardized reporting unit, differences in positivity rates remain[21]. Policy makers should be cautious of these differences in performance and not rely on previous research solely. When implementing a screening program a pilot-phase, encompassing all screening ages, should be conducted prior to assess most optimal cut-off resulting in an acceptable number of colonoscopy referrals with an sufficient diagnostic yield.

In the Netherlands, screening is done by a quantitative FIT. Both quantitative FITs as well as qualitative FITs are available for CRC-screening, though qualitative tests are infrequently used, and seem to be less sensitive than the quantitative tests. Quantitative tests can specifically measure

fecal Hb in µg Hb per gram feces. Despite this quantitative nature of the test, FITs are invariably used in a qualitative manner (i.e. using a pre-specified cut-off and reporting only a positive or negative result). By using FIT as a qualitative test, a great window of opportunity for personalized screening strategies is lost. Fecal Hb concentration can be used in various manners, including:

- using fecal Hb concentrations to predict an individuals risk of having advanced neoplasia, based on both risk factors as well as FIT-results.
- using fecal Hb concentrations of even very low values (i.e. under most currently used threshold) over multiple rounds to refer people for colonoscopy or to extend screening intervals.
- using different fecal Hb concentrations for men and women, thereby also incorporating different age categories.
- using fecal Hb concentrations to predict who is at risk of advanced neoplasia, and more specifically of large advanced adenomas. A pre-selection of these patients could, in theory, prevent the need for a second look colonoscopy.
- using fecal Hb concentrations after a colonoscopy to determine colonoscopy surveillance intervals.

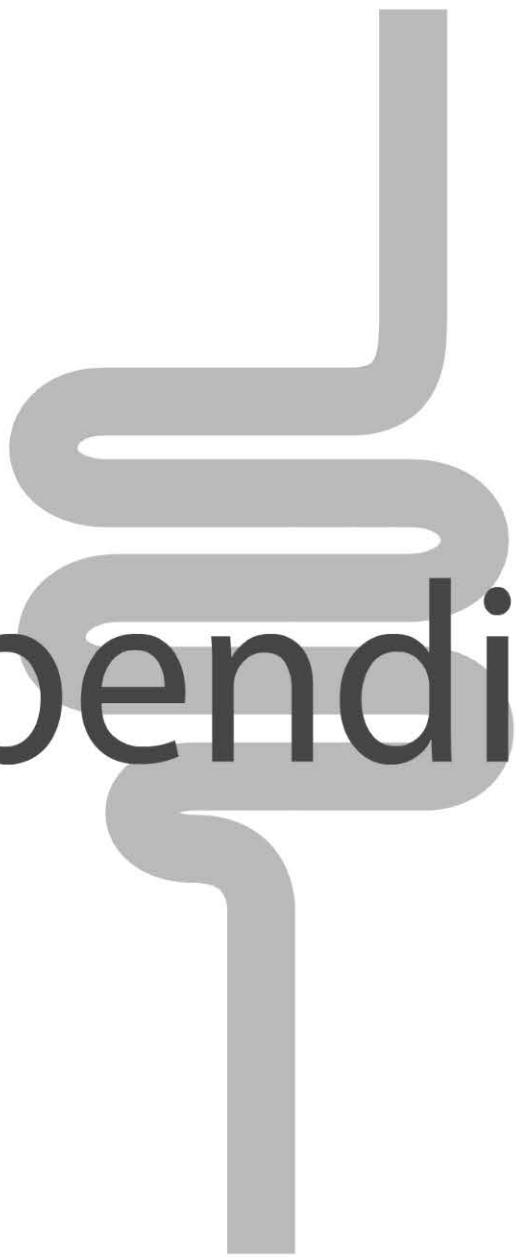
While some of the above-mentioned examples have been already explored in our cohort, other scenarios still require further determination, for example using MISCAN modeling. Also, the usability of such scenarios needs to be confirmed in other large screening settings.

In the near future, the first results on long-term FIT screening and colonoscopy screening will be available. These results will show whether the expected decrease in CRC incidence, due to the removal of pre-cancerous adenomas, will really be overt on a population level. With respect to the most optimal screening method, until more effective, non-invasive and practical screening methods arrive, FIT remains the screening strategy of choice. However, there is still much to gain regarding the use of fHb by quantitative FIT results in personalized-screening strategies.

References

1. Kuipers EJ, Grady WM, Lieberman D, et al. Colorectal cancer. Nature Reviews Disease Primers 2015: article number 15065. doi:10.1038/nrdp.2015.65
1. Schreuders EH, Ruco A, Rabeneck L, et al. Colorectal cancer screening: a global overview of existing programmes. Gut 2015; 64(10): 1637-49.
2. Winawer SJ, Zauber AG, Fletcher RH, et al. Guidelines for colonoscopy surveillance after polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer and the American Cancer Society. CA Cancer J Clin 2006; 56(3): 143-59; quiz 84-5.
3. Gellad ZF, Stechuchak KM, Fisher DA, et al. Longitudinal adherence to fecal occult blood testing impacts colorectal cancer screening quality. Am J Gastroenterol 2011; 106(6): 1125-34.
4. Fraser CG, Allison JE, Halloran SP, Young GP, Org WE. A Proposal to Standardize Reporting Units for Fecal Immunochemical Tests for Hemoglobin. Jnci-J Natl Cancer I 2012; 104(11): 810-4.
5. Digby J, Fraser CG, Carey FA, Lang J, Stanners G, Steele RJ. Interval cancers using a quantitative faecal immunochemical test (FIT) for haemoglobin when colonoscopy capacity is limited. J Med Screen 2015.
6. Steele RJ, McDonald PJ, Digby J, et al. Clinical outcomes using a faecal immunochemical test for haemoglobin as a first-line test in a national programme constrained by colonoscopy capacity. United European Gastroenterol J 2013; 1(3): 198-205.
7. Essink-Bot ML, Dekker E. Equal access to colorectal cancer screening. Lancet 2016; 387(10020): 724-6.
8. Lee JK, Liles EG, Bent S, Levin TR, Corley DA. Accuracy of fecal immunochemical tests for colorectal cancer: systematic review and meta-analysis. Ann Intern Med 2014; 160(3): 171.
9. Lieberman D. Screening for colorectal cancer in individuals at average risk: current methods and emerging issues. JAMA Intern Med 2014; 174(1): 10-1.
10. de Wijkerslooth TR, Stoop EM, Bossuyt PM, et al. Immunochemical fecal occult blood testing is equally sensitive for proximal and distal advanced neoplasia. Am J Gastroenterol 2012; 107(10): 1570-8.
11. Regula J, Rupinski M, Kraszewska E, et al. Colonoscopy in colorectal-cancer screening for detection of advanced neoplasia. N Engl J Med 2006; 355(18): 1863-72.
12. Brenner H, Hoffmeister M, Stegmaier C, Brenner G, Altenhofen L, Haug U. Risk of progression of advanced adenomas to colorectal cancer by age and sex: estimates based on 840,149 screening colonoscopies. Gut 2007; 56(11): 1585-9.
13. Ferlitsch M, Reinhart K, Pramhas S, et al. Sex-specific prevalence of adenomas, advanced adenomas, and colorectal cancer in individuals undergoing screening colonoscopy. JAMA 2011; 306(12): 1352-8.
14. eMarketer. Worldwide Smartphone Usage to Grow 25% in 2014. http://wwwemarketercom/Article/Worldwide-Smartphone-Usage-Grow-25-2014/1010920 2014.
15. Austoker J, Giordano L, Hewitson P, Villain P, International Agency for Research on C. European guidelines for quality assurance in colorectal cancer screening and diagnosis. First Edition--Communication. Endoscopy 2012; 44 Suppl 3: SE164-85.
16. Rex DK, Bond JH, Winawer S, et al. Quality in the technical performance of colonoscopy and the continuous quality improvement process for colonoscopy: recommendations of the U.S. Multi-Society Task Force on Colorectal Cancer. Am J Gastroenterol 2002; 97(6): 1296-308.
17. Kaminski MF, Regula J, Kraszewska E, et al. Quality indicators for colonoscopy and the risk of interval cancer. N Engl J Med 2010; 362(19): 1795-803.

18. Corley DA, Jensen CD, Marks AR, et al. Adenoma detection rate and risk of colorectal cancer and death. N Engl J Med 2014; 370(14): 1298-306.
19. Grady D. Report Affirms Lifesaving Role of Colonoscopy. The New York Times; 2012.
20. Chiang TH, Chuang SL, Chen SL, 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.



appendices

Introductie (Deel I)

Dikke darmkanker is een veelvoorkomende ziekte. In Nederland worden jaarlijks ongeveer 12.000 mensen gediagnosticeerd met darmkanker[1]. Via een bevolkingsonderzoek kunnen mensen gescreend worden op darmkanker zodat de ziekte in een vroeg stadium wordt ontdekt of in een voorstadium van de ziekte. Wanneer dikke darmkanker in een vroeg stadium wordt vastgesteld is de kans op genezing groter[2]. Darmkanker begint als een poliep, dit is een verdikking in de darm. Veel poliepen worden nooit kwaadaardig, maar een klein deel groeit verder en ontwikkelt zich uiteindelijk tot darmkanker. Door het verwijderen van deze zogenoemde hoog-risico poliepen, het voorstadium van darmkanker, kan dikke darmkanker worden vóórkomen. Er zijn verschillende manier om te screenen op darmkanker. Dit proefschrift richt zich op ontwikkelingen op het gebied van dikke darmkankerscreening en de rol hierin van de fecale immunochemische test (FIT). In deel I van dit proefschrift word een uitgebreid overzicht gegeven over de huidige status van dikke darmkankerscreening en de verschillende beschikbare screeningsmethoden.

Darmkankerscreeningsmethoden met een focus op FIT (Deel II)

Om te screenen naar darmkanker zijn verschillende methoden beschikbaar en deze kunnen grofweg in twee strategieën worden verdeeld; beeldvorming van de dikke darm (i.e. colon) met invasieve procedures zoals een kijkonderzoek van de darm (i.e. coloscopie), en niet-invasieve strategieën, zoals de fecale occult bloed test (FOBT)[2]. Deze laatste test, de FOBT, meet of er bloed aanwezig is in de ontlasting. Bloed in de ontlasting kan veroorzaakt worden door poliepen of kanker in de dikke darm, maar dit kan ook andere oorzaken hebben. Indien er bloed wordt gevonden in de ontlasting wordt de patiënt doorverwezen voor vervolgonderzoek. Voorheen werd voor het testen met de FOBT vaak de guaiac FOBT (gFOBT) gebruikt, echter, op dit moment wordt de gFOBT veelal vervangen door de FIT[3]. Om deze twee testen te vergelijken hebben we een systematisch review en meta-analyse uitgevoerd waarin de testkarakteristieken en nauwkeurigheid van de FIT en gFOBT werden vergeleken **(hoofdstuk 2)**. De resultaten van dit onderzoek toonden een hogere gevoeligheid voor zowel darmkanker als voor de hoog-risico poliepen bij het screenen met de FIT in vergelijking tot de gFOBT.

Screenen voor darmkanker kan door mensen direct een coloscopie aan te bieden (primaire coloscopiescreening), door een deel van de dikke darm te onderzoeken met een kijkonderzoek (primaire sigmoidoscopiescreenign) of door eerst een FOBT aan te bieden. Een sigmoidoscopie

is eenzelfde onderzoek als de coloscopie, alleen wordt bij deze procedure niet de gehele darm bekeken, maar alleen de endeldarm en het rectum. Indien er afwijkingen worden gevonden bij de sigmoidoscopie wordt de patiënt verwezen voor een coloscopie. Voor FOBT-screening geld dat indien er bloed in de ontlasting wordt gevonden bij de FOBT, de patiënt doorverwezen wordt voor coloscopie. Coloscopie wordt beschouwd als de gouden standaard voor de diagnose van dikke darmkanker. Indien er geen afwijkingen worden gevonden bij coloscopie, is het advies om de coloscopie na 10 jaar te herhalen[4]. Indien er echter geen bloed wordt aangetoond bij een FOBT (i.e., gFOBT of FIT) dan wordt aanbevolen om de test jaarlijks of elke twee jaar te herhalen[5]. Tot op heden hebben de meeste wetenschappelijke onderzoeken zich gericht op het vergelijken van één ronde primair coloscopiescreening versus één ronde FIT-screening (10 versus 2 jaar) en zijn er geen studies bekend die coloscopiescreening met meerdere rondes FIT vergelijken. Om deze reden hebben we coloscopiescreening, flexibele sigmoïdoscopie (FS) screening en vier rondes FIT-screening met elkaar vergeleken **(hoofdstuk 3)**. De resultaten van het vergelijkend onderzoek lieten een betere prestatie voor de FIT-screening zien, waarbij tweemaal meer hoog-risico poliepen werden gevonden bij FIT dan bij screening met FS of coloscopie. Hoog-risico poliepen werden gevonden in 4,5% (95% CI 4,2-4,9) vs. 2,3% (95% CI 2,0-2,7) en 2,2% (95% CI 1,8-2,6) respectievelijk onder alle genodigden. Bovendien werd bij FIT-screening driemaal vaker dikke darmkanker gevonden (0,6% tegenover 0,2%; p <0,001). Toch dient te worden opgemerkt dat deze resultaten met voorzichtigheid moeten worden geïnterpreteerd. Ten eerste worden deze resultaten sterk beïnvloed door het aantal deelnemers aan de screeningstest. Ten tweede zou het zo kunnen zijn dat hoog-risico poliepen die verwijderd zijn bij coloscopie mogelijk in de toekomst dikke darmkanker hadden kunnen worden. Dit laatste gegeven maakt dat er mogelijk een onderschatting is van het aantal gevonden en voorkomen darmkankers bij coloscopiescreening in vergelijking met FIT-screening.

In 2014 werd een landelijk bevolkingsonderzoek naar dikke darmkanker in Nederland gestart. Door middel van een nationale aanbesteding werd besloten om te screenen een FIT van het merk FOB-Gold (Sentinel, Italië). Echter, de Nederlandse wetenschappelijke onderzoeken voorafgaand aan het landelijk bevolkingsonderzoek zijn gedaan met een ander merk, de OC-sensor (Eiken, Japan). Om te onderzoeken hoe deze twee tests zich tot elkaar verhouden, hebben we een gerandomiseerde trial uitgevoerd in de vierde ronde van het proefbevolkingsonderzoek **(hoofdstuk 4)**. Dit proefbevolkingsonderzoek bestaat sinds 2006 en betreft een geselecteerde groep mensen die elke 2 jaar voor darmkankerscreening met FIT zijn uitgenodigd. De resultaten in de vierde ronde lieten zien, dat beide tests dezelfde positieve voorspellende waarde hadden. Echter, er waren tevens meer positieve testen bij de OC-sensor in vergelijking met de FOB-Gold (7,9% vs. 6.5%; p = 0,002). Dit

hogere aantal positieve testen leidde vervolgens tot een hoger verwijscijfer voor coloscopie en zodoende tot een hogere diagnostische opbrengst. Opvallend was dat het hogere aantal positieve testen werd gevonden, ondanks dat voor beide testen eenzelfde afkapwaarde was gebruikt. Onze resultaten laten zelfs zien dat bij het gebruik van een gestandaardiseerde maat voor de afkapwaarde er verschillen optraden met betrekking tot het aantal positieve testen. Het aantal positieve testen is een belangrijk gegeven, omdat dit een directe weerspiegeling van het benodigde aantal coloscopieën is. Momenteel hebben veel landen wachttijden door de hoge aantallen coloscopieën die nodig zijn voor een bevolkingsonderzoek naar dikke darmkanker[6,7].

Bij het screenen naar dikke darmkanker middels FIT is een stabiele en hoge deelname over meerdere rondes belangrijk. We toonden aan dat de deelname aan screening stabiel was over ten minste drie screeningsrondes , maar dat het aantal positieve testen en de positief voorspellende waarde daalden **(hoofdstuk 5)**. Het percentage positieve testen was het hoogst in de eerste ronde (8,4%) en nam af in de loop van de tweede (6,0%) en derde (5,7%) screening ronde ($p < 0,001$). Logischerwijs nam ook de diagnostische opbrengst af over de rondes (ronde 1: 3,3%; ronde 2: 1,9%, en ronde 3: 1,3%; $p < 0,001$).

In het volgende hoofdstuk **(hoofdstuk 6)** werd de deelname over meerdere rondes uitgebreider beschreven, waarbij werd aangetoond dat deelname aan dikke darmkanker stabiel bleef gedurende vier screeningsrondes en dat de deelname zelfs licht toenam. Over alle vier de rondes, deed 72% ten minste één keer mee met screening, terwijl 28% van de uitgenodigde mensen in geen enkele ronde deelnam. Deze laatste groep bestond voornamelijk uit mannen met een lage sociaaleconomische status. In toekomstig onderzoek is er meer aandacht voor deze groep nodig omdat dit ook als een hoog-risico groep beschouwd wordt voor het ontwikkelen van darmkanker. Derhalve is vervolgonderzoek nodig om nieuwe strategieën te onderzoeken om deelname aan screening in deze groep te verhogen. Momenteel krijgt iedereen eenzelfde uitnodiging voor darmkanker screening, terwijl het mogelijk juist noodzakelijk is dat uitnodigingsstrategieën verschillen tussen subgroepen om zo gelijke kansen op optimale resultaten te bereiken[8].

Ondanks de hoge deelname aan FIT-screening en een relatief hoge gevoeligheid van de FIT voor dikke darmkanker van 75-85%, is de gevoeligheid van de FIT voor de detectie van hoog-risico poliepen veel lager. Om deze reden is het belangrijk om onderzoek te doen of de gevoeligheid van de FIT verbeterd kan worden. Hiervoor hebben we onderzocht of het combineren van de FIT met het meten van dikke darmbacteriën (i.e. het microbioom) een verbetering in de diagnose van

darmkanker zou zijn **(hoofdstuk 7).** Voor dit onderzoek werden vier bacteriën gemeten waarvan bekend is dat deze mogelijk een rol spelen in de ontwikkeling van darmkanker (Escherichia coli, Fusobacterium nucleatum, Bacteroides en Fecalibacterium prausnitzii). Onze resultaten toonden geen toegevoegde waarde van het combineren van de FIT en het microbioom wanneer we naar elke bacterie apart keken, maar we vonden wel dat het totale aantal bacteriën hoger was in de ontlasting van patiënten met hoog-risico poliepen en dikke darmkanker.

Het gebruik van FIT als een kwantitatieve gids binnen de darmkankerscreening (Deel III)

Een FIT kan zowel kwalitatief zijn, wat leidt tot een positief of negatief resultaat, of kwantitatief waarbij de specifieke hoeveelheid bloed in de ontlasting kan worden gemeten[3]. Hoewel de kwantitatieve FIT vaakst wordt gebruik bij bevolkingsonderzoeken, worden de resultaten steevast op een kwalitatieve manier gerapporteerd. Dit houdt in dat de specifieke bloedwaarde niet wordt gerapporteerd, maar er alleen een positief of een negatief resultaat uit de test komt welke gebaseerd is op een vooraf ingestelde afkapwaarde. In dit derde deel van het proefschrift onderzoeken we of de specifieke bloedwaarde (i.e., fecaal bloed) in de ontlasting gebruikt kan worden om risico's te schatten en screeningstrategieën te verbeteren.

Ten eerste hebben we de concentratie fecaal bloed van patiënten onderzocht. Hierbij hebben we gekeken of de concentratie onder de afkapwaarde gebruikt kan worden om in te schatten wie het hoogste risico heeft op poliepen of darmkanker. We toonden aan het fecaal bloed van een FIT onder de afkapwaarde kan voorspellen wie in de komende jaren het meeste risico heeft op het ontwikkelen van hoog-risico poliepen en darmkanker **(Hoofdstuk 8)**. Deelnemers met een fecale bloedconcentratie van > 8 µg Hb/g hadden een achtvoudig hoger risico op het ontwikkelen van hoog-risico poliepen en darmkanker, in vergelijking met mensen met een fecaal bloedconcentratie van 0 µg Hb/g. Bovendien lieten onze resultaten zien dat opeenvolgende FIT resultaten nog meer inzicht in deze risico's geven en te gebruiken zijn om gepersonaliseerde strategieën op te baseren. Ook lieten de resultaten zien dat mannen een twee keer zo hoog risico hebben op het ontwikkelen van darmkanker als vrouwen.

Ondanks bekende verschillen tussen mannen en vrouwen, worden zij steevast gescreend op dezelfde manier[9,10]. In **hoofdstuk 9** gaven we een overzicht van deze verschillen , in een screeningscohort waarbij alle deelnemers zowel een FIT hebben gedaan als een coloscopie hebben ondergaan. Met behulp van deze gegevens werkten we twee screeningsstrategieën uit: 1) een strategie waarbij de

sensitiviteit van de FIT gelijk is voor mannen en vrouwen kunnen beide niveaus gevoeligheid bij mannen en vrouwen (i.e., een lagere afkapwaarde bij vrouwen); 2) een strategie waarbij het aantal gemiste laesies gelijk is voor mannen en vrouwen (i.e., een lagere afkapwaarde bij mannen). Beide strategieën kunnen gebruikt worden, afhankelijk van het beoogde resultaat: een gelijke sensitiviteit versus een gelijk aantal gemiste leasies voor mannen en vrouwen.

Vervolgens hebben we gekeken naar fecale bloedconcentraties van patiënten met een concentratie boven de afkapwaarde. Deze resultaten lieten zien dat fecaal bloed gebruikt kan worden om te voorspellen wie een hoog risico heeft om een tweede coloscopie binnen 1 jaar moet ondergaan **(hoofdstuk 10)**. Daarnaast bleek dat binnen deze patiëntenpopulatie er relatief vaak een tweede coloscopie moet plaatsvinden, en dat dit uiteindelijk tot 12% extra coloscopie capaciteit vergt binnen een screeningsprogramma. Tenslotte hebben we aangetoond dat fecale bloedconcentraties kunnen worden gebruikt screeningsintervallen op te baseren en om zo screeningprogramma's te optimaliseren **(hoofdstuk 11)**.

Kwaliteit en endoscopie binnen darmkankerscreening (Deel IV)

In deel IV werden kwaliteitszaken rondom darmkankerscreening besproken. Ten aanzien van de kwaliteit van patiënteninformatie, hebben we laten zien dat er nog veel te winnen valt bij het verstrekken van volledige en begrijpelijke informatie voor degenen die overwegen deel te nemen aan dikke darmkankerscreening **(hoofdstuk 12)**. We hebben een website nauwkeurigheid score (WAS) ontwikkeld en hiermee samen met gevalideerde scoresystemen medische websites beoordeeld op kwaliteit en volledigheid van informatie over darmkankerscreening. Opvallend was er een slechte overlap was tussen de kwaliteit van websites en hun plaats binnen Google. Deze resultaten suggereren dat mensen belangrijke en kwalitatief goede websites zullen missen, omdat deze websites pas onderaan de resultaten van Google staan. Er moet om deze reden meer aandacht worden besteed aan de kwaliteit en begrijpelijkheid van medische informatie op het internet, omdat gebruik van internet voor medische informatie in de toekomst zal blijven toenemen.

Bij een bevolkingsonderzoek is het aantal gemiste darmkankers net zo belangrijk als het aantal gevonden darmkankers. Om deze reden hebben we het aantal gevonden en gemiste kankers binnen vier screeningsrondes met FIT vergeleken; in 3,9% van de FIT-positieve screenees werd darmkanker gevonden, en in 0,14% van de FIT-negatieve screenees werd een darmkanker gemist voordat de volgende FIT test gepland stond (**hoofdstuk 13**). In 0,43% van de patiënten werd darmkanker

gediagnosticeerd binnen 10 jaar na de coloscopie. Dikke darmkankers die gevonden werden bij het screenen hadden een gunstiger stadium dan darmkankers die gediagnosticeerd werden buiten het screeningsprogramma. Over meerdere screeningsrondes neemt het aantal kankers dat in een vroeg stadium wordt ontdekt toe. Helaas worden deze laesies vaak onvoldoende herkend bij coloscopie. Dit is een belangrijke kwestie, omdat herkenning leidt tot adequate resectie van de kanker en zo een onnodige operatie kan voorkomen. We toonden aan dat in ons proefbevolkingsonderzoek cohort, meer dan de helft vroege darmkankers niet als zodanig werden geïdentificeerd bij eerste coloscopie **(hoofdstuk 14)**. Daarom moet er in de toekomst meer aandacht komen voor het diagnosticeren en verwijderen van deze laesies.

Kwaliteit van een screeningsprogramma is sterk afhankelijk van de kwaliteit van coloscopie. Om deze reden hebben we de kwaliteit van de coloscopie tussen ziekenhuizen gemeten. Deze resultaten lieten duidelijke verschillen tussen ziekenhuizen zien **(hoofdstuk 15)**. Toekomstig onderzoek zou zich moeten richten op deze verschillen tussen ziekenhuizen om zo factoren te identificeren om de kwaliteit van coloscopiën te verbeteren.

FIT naar de toekomst

Onderzoek naar darmkankerscreening heeft een enorme ontwikkeling doorgemaakt in de afgelopen decennia. Hoewel verbeteringen zijn aangebracht met betrekking tot de kwaliteit en de nauwkeurigheid van screening strategieën, zijn de strategieën zelf niet veel veranderd en bestaan deze nog steeds voornamelijk uit een coloscopiescreening of FOBT screening. De perfecte screeningstest (met 100% sensitiviteit en specificiteit, niet-invasieve en een test waaraan iedereen mee wil doen) bestaat helaas nog niet en het is onwaarschijnlijk dat een dergelijke test beschikbaar zal zijn in de komende jaren.

In Nederland is er sinds januari 2014 een bevolkingsonderzoek naar darmkanker met FIT. Op basis van eerder onderzoek, werd een screeningsinterval van 2 jaar gekozen. Momenteel is implementatie van het bevolkingsonderzoek nog bezig en worden gefaseerd alle inwoners in de leeftijd tussen 55 en 74 jaar uitgenodigd voor het bevolkingsonderzoek.

In de nabije toekomst zullen de eerste resultaten op de lange termijn FIT screening en coloscopiescreening beschikbaar komen. Deze resultaten zullen aantonen of de verwachte daling van dikke darmkankerincidentie zich ook daadwerkelijk zo zal ontwikkelen op populatieniveau.

Met betrekking tot de meest optimale screeningmethode blijft FIT de eerste keuze, zeker gezien de hoge deelnamegraad en praktische voordelen. Voor de toekomst valt er nog veel te winnen met betrekking tot het gebruik van patiëntkarakteristieken en fecale bloedconcentraties in gepersonaliseerde-screening strategieën.

Referenties

1. Nederlandse Kankerregistratie. 2012. www.cijfersoverkanker.nl2013).
2. Kuipers EJ, Grady WM, Lieberman D, et al. Colorectal cancer. Nature Reviews Disease Primers 2015; article number 15065. doi:10.1038/nrdp.2015.6.
3. Schreuders EH, Ruco A, Rabeneck L, et al. Colorectal cancer screening: a global overview of existing programmes. Gut 2015; 64(10): 1637-49.
4. Lieberman DA, Rex DK, Winawer SJ, et al. Guidelines for colonoscopy surveillance after screening and polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer. Gastroenterology 2012; 143(3): 844-57.
5. 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.
6. Digby J, Fraser CG, Carey FA, Lang J, Stanners G, Steele RJ. Interval cancers using a quantitative faecal immunochemical test (FIT) for haemoglobin when colonoscopy capacity is limited. J Med Screen 2015.
7. Steele RJ, McDonald PJ, Digby J, et al. Clinical outcomes using a faecal immunochemical test for haemoglobin as a first-line test in a national programme constrained by colonoscopy capacity. United European Gastroenterol J 2013; 1(3): 198-205.
8. Essink-Bot ML, Dekker E. Equal access to colorectal cancer screening. Lancet 2016; 387(10020): 724-6.
9. de Wijkerslooth TR, Stoop EM, Bossuyt PM, et al. Immunochemical fecal occult blood testing is equally sensitive for proximal and distal advanced neoplasia. Am J Gastroenterol 2012; 107(10): 1570-8.
10. Brenner H, Hoffmeister M, Stegmaier C, Brenner G, Altenhofen L, Haug U. Risk of progression of advanced adenomas to colorectal cancer by age and sex: estimates based on 840,149 screening colonoscopies. Gut 2007; 56(11): 1585-

Abbreviations

AA	advanced adenoma
ADR	adenoma detection rate
AN	advanced neoplasia
ASA	American society of anesthesiologists
ASR	age standardized ratio
BBPS	Boston bowel preparation score
BCO	Below the cut-off
BMI	body mass index
CI	confidence interval
CIR	cecal intubation rate
CQI	colonoscopy quality indicator
CRC	colorectal cancer
CTC	computed tomography colonography
DTA	diagnostic test accuracy
FIT	fecal immunochemical test
fHb	fecal hemoglobin
FOBT	fecal occult blood test
gFOBT	guaiac fecal occult blood test
FN	false negatives
FP	false positives
FS	flexible sigmoidoscopy
HGD	high grade dysplasia
HP	hyperplastic polyp
HR	hazard ratio
IQR	inter quartile range
MAP	mean adenomas per procedure
OR	odds ratio
PC	post-colonoscopy
PR	Positivity rate
SD	standard deviation
SP	serrated polyp
SSA	sessile serrated adenoma

TN	true negatives
TP	true postives
TVA	(tubulo) villous adenoma
TA	tubular adenoma
UICC	union internationale centre le cancer

Contributing authors

Listed in alphabetical order.
Affiliations at time that research was conducted.

Marjolein van Ballegooijen
Department of Public Health
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Tim D.G. Belderbos
Department of Gastroenterology and Hepatology
University Medical Center Utrecht
Utrecht, The Netherlands

Blerdi Blakaj
Department of Gastroenterology and Hepatology
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Evelien .J.C. Bongers
Foundation of Population Screening Mid-West
Amsterdam, The Netherlands

Frank ter Borg
Department of Gastroenterology and Hepatology
Deventer Hospital
Deventer, The Netherlands

Pieter C.J. ter Borg
Department of Gastroenterology and Hepatology
Ikazia Hospital
Rotterdam, The Netherlands

Gerard J.J. M. Borsboom
Department of Public Health
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Amanda Bos
Department of Research
Netherlands Comprehensive Cancer Organisation (IKNL)
Utrecht, The Netherlands

Patrick M. Bossuyt
Department of Clinical Epidemiology, Biostatistics and Bioinformatics
Academic Medical Centre Amsterdam
Amsterdam, The Netherlands

Marco J. Bruno
Department of Gastroenterology and Hepatology
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Leonie van Dam
Department of Gastroenterology and Hepatology
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Evelien Dekker
Department of Gastroenterology and Hepatology
Academic Medical Centre Amsterdam
Amsterdam, The Netherlands

Marie-Louise Essink-Bot†
Department of Publich Health
Academic Medical Centre Amsterdam
Amsterdam, The Netherlands

Doris M. Le Fèvre
Achmea Health Care
Leiden, The Netherlands

Gwenny M. Fuhler
Department of Gastroenterology and Hepatology
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Onno van der Galiën
Achmea Health Care
Leiden, The Netherlands

Bettina E. Hansen
Department of Gastroenterology and Hepatology
Department of Public Health
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Ulrike Haug
Department of Clinical Epidemiology
Leibniz Institute for Prevention Research and Epidemiology – BIPS
Bremen, Germany

Theo J. Hiemstra
Achmea Health Care
Leiden, The Netherlands

Lieke Hol
Department of Gastroenterology and Hepatology
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Kirsten Izelaar
Regional organization for Population Screening South-West Netherlands
Rotterdam, The Netherlands

Atija Kapidzic
Department of Gastroenterology and Hepatology
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Arjun Koch
Department of Gastroenterology and Hepatology
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Sergey R. Konstantinov
Department of Gastroenterology and Hepatology
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Ernst J. Kuipers
Department of Gastroenterology and Hepatology
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Iris Lansdorp-Vogelaar
Department of Public Health
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Monique E. van Leerdam
Gastroenterology and Hepatology
Netherlands Cancer Institute
Amsterdam, The Netherlands

Rosalie C. Mallant-Hent
Department of Gastroenterology and Hepatology
Flevoziekenhuis
Almere, The Netherlands
Maarten A.C. Meijssen
Department of Gastroenterology and Hepatology
Isala Hospital
Zwolle, The Netherlands

Leon M.G. Moons
Department of Gastroenterology and Hepatology
University Medical Center Utrecht
Utrecht, The Netherlands

Marco W. Mundt
Department of Gastroenterology and Hepatology
Flevoziekenhuis
Almere, The Netherlands

Martijn G.H. van Oijen
Department of Gastroenterology and Hepatology
University Medical Center Utrecht
Utrecht, The Netherlands

Rob J.T.H. Ouwendijk
Department of Gastroenterology and Hepatology
Ikazia Hospital
Rotterdam, The Netherlands

Maikel P. Peppelenbosch
Department of Gastroenterology and Hepatology
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Aafke H.C. van Roon
Department of Gastroenterology and Hepatology
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Peter van der Schaar
Department of Gastroenterology and Hepatology
St. Antonius Hospital
Nieuwegein, The Netherlands

Eline H. Schreuders
Department of Gastroenterology and Hepatology
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Peter D. Siersema
Department of Gastroenterology and Hepatology
University Medical Center Utrecht
Utrecht, The Netherlands

Manon C.W. Spaander
Department of Gastroenterology and Hepatology
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Wolfert J. Spijker
Regional organization for Population Screening South-West Netherlands
Rotterdam, The Netherlands

Ewout W. Steyerberg
Department of Public Health
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Esther M. Stoop
Department of Gastroenterology and Hepatology
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Merijn T. Stouten
Gupta Strategists
Ophemert, The Netherlands

An K. Stroobants
Department of Clinical Chemistry
Academic Medical Centre Amsterdam
Amsterdam, The Netherlands

Thjon J. Tang
Department of Gastroenterology and Hepatology
IJsselland Hospital
Capelle aan de IJssel, The Netherlands

Sander J.O. Veldhuyzen van Zanten
Division of Gastroenterology
University of Alberta Hospital
Edmonton, Alberta, Canada

Manon van der Vlugt
Department of Gastroenterology and Hepatology
Academic Medical Centre Amsterdam
Amsterdam, The Netherlands

Wouter H. de Vos tot Nederveen Cappel
Department of Gastroenterology and Hepatology
Isala Hospital
Zwolle, The Netherlands

Anneke .J. van Vuuren

Department of Gastroenterology and Hepatology

Erasmus MC University Medical Center

Rotterdam, The Netherlands

Els Wieten

Department of Gastroenterology and Hepatology

Erasmus MC University Medical Center

Rotterdam, The Netherlands

Thomas R. de Wijkerslooth

Department of Gastroenterology and Hepatology

Academic Medical Centre Amsterdam

Amsterdam, The Netherlands

Ann G. Zauber

Epidemiology and Biostatistics

Memorial Sloan Kettering Cancer Center

New York, United States of America

Bibliography

1. **Esmée J. Grobbee**, Manon van der Vlugt, Anneke J. van Vuuren, An K. Stroobants, Marco W. Mundt, Wolfert J. Spijker, Evelien J.C. Bongers, Ernst J. Kuipers, Iris Lansdorp-Vogelaar, Patrick M. Bossuyt, Evelien Dekker, Manon C.W. Spaander. A randomised comparison of two fecal immunochemical tests in population-based colorectal cancer screening, Gut 2016 Aug 9. pii: gutjnl-2016-311819. doi: 10.1136/gutjnl-2016-311819. [Epub ahead of print]

2. **Esmée J. Grobbee*,** Eline H. Schreuders*, Aafke H.C. van Roon, Leonie van Dam, Ann G. Zauber, I. Lansdorp-Vogelaar, Gerard J.J. M. Borsboom, Ewout W. Steyerberg, Monique E. van Leerdam, Manon C.W. Spaander, Ernst J. Kuipers. *both authors contributed equally. Guaiac-based faecal occult blood tests versus fecal immunochemical tests for colorectal cancer screening in average-risk individuals, manuscript submitted at Cochrane Diagnostic Test Accuracy Library.

3. **Esmée J. Grobbee*,** Manon van der Vlugt*, Anneke J. van Vuuren, An K. Stroobants, Rosalie C. Mallant-Hent, Iris Lansdorp-Vogelaar, Patrick M. Bossuyt, Ernst J. Kuipers, Evelien Dekker, Manon C.W. Spaander *both authors contributed equally. Prospective comparison of three different colorectal cancer screening strategies; colonoscopy, flexible sigmoidoscopy and multiple rounds of FIT, manuscript submitted.

4. **Esmée J. Grobbee,** Gwenny M. Fuhler, Blerdi Blakaj, Sergey R. Konstantinov, Marco J. Bruno, Maikel P. Peppelenbosch, Ernst J. Kuipers, Manon C.W. Spaander. First steps towards combining fecal immunochemical testing with the gut microbiome in colorectal cancer screening, manuscript submitted.

5. **Esmée J. Grobbee,** Eline H. Schreuders, Bettine E. Hansen, Iris Lansdorp-Vogelaar, Marco J. Bruno, Manon C.W. Spaander, Ernst J. Kuipers . Fecal haemoglobin concentrations predict future advanced colorectal neoplasia in long-term population-based FIT-screening, manuscript submitted.

6. **Esmée J. Grobbee,** Els Wieten, Bettina E. Hansen, Esther M. Stoop, Thomas R. de Wijkerslooth, Iris Lansdorp-Vogelaar, Patrick M. Bossuyt, Evelien Dekker, Ernst J. Kuipers, Manon C.W. Spaander. Fecal immunochemical test-based colorectal cancer screening: the gender

dilemma, United European Gastroenterology Journal [E-pub ahead of print].

7. **Esmée J. Grobbee**, Atija Kapidzic, Anneke J. van Vuuren, Monique van Leerdam, Iris Lansdorp-Vogelaar, Caspar W.N. Looman, Marco J. Bruno, Ernst J. Kuipers, Manon C.W. Spaander. Second look colonoscopies and the impact on capacity in FIT-based colorectal cancer screening, American Journal of Gastroenterol. 2015 Jul;110(7):1072-7.

8. **Esmée J. Grobbee***, Manon van der Vlugt*, Patrick M. Bossuyt, Amanda Bos, Evelien Bongers, Wolfert Spijker, Iris Lansdorp-Vogelaar, Ernst J. Kuipers, Manon C.W. Spaander, Evelien Dekker *both authors contributed equally. Screen-detected and non-screen-detected colorectal cancers after four rounds of fecal immunochemical test-based colorectal cancer screening. Manuscript in preparation.

9. **Esmée J. Grobbee*,** Tim D.G. Belderbos*, Martijn G.H. van Oijen, Maarten A.C. Meijssen, Rob J.T.H. Ouwendijk, Thjon J. Tang, Frank ter Borg, Peter van der Schaar, Doris M. Le Fèvre, Merijn T. Stouten, Onno van der Galiën, Theo J. Hiemstra, Wouter H. de Vos tot Nederveen Cappel, Pieter C.J. ter Borg, Manon C.W. Spaander, Leon M.G. Moons, Ernst J. Kuipers, Peter D. Siersema*both authors contributed equally. Comparison of cecal intubation and adenoma detection between hospitals can provide incentives to improve quality of colonoscopy . Endoscopy. 2015 Aug;47(8):703-9.

10. Atija Kapidzic, **Esmée J. Grobbee**, Liek Hol, Aafke H. van Roon, Anneke J. van Vuuren, Wolfert Spijker, Kirsten Izelaar, Marjolein van Ballegooijen, Ernst J. Kuipers, Monique E. van Leerdam. Attendance and yield over three rounds of population-based fecal immunochemical test screening, American Journal of Gastroenterology. 2014 Aug;109(8):1257-64.

11. Manon van der Vlugt, **Esmée J. Grobbee**, Patrick M.M. Bossuyt, Evelien Bongers, Wolfert Spijker, Ernst J. Kuipers, Iris Lansdorp-Vogelaar, Marie-Louise Essink-Bot†, Manon C.W. Spaander, Evelien Dekker. Adherence to colorectal cancer screening: 4 rounds of fecal immunochemical test (FIT) based screening. Manuscript submitted.

12. Ulrike Haug, **Esmée J. Grobbee**, Iris Lansdorp-Vogelaar, Manon C.W. Spaander, Ernst Kuipers. Immunochemical fecal occult blood testing to screen for colorectal cancer: can the screening interval be extended? Gut. 2016 Mar 22.. doi: 10.1136 [Epub ahead of print].

13. Eline H. Schreuders, **Esmée J. Grobbee**, Ernst J. Kuipers, Manon C.W. Spaander, Sander J.O. Veldhuyzen van Zanten. Systematic Assessment of the Quality of Patientoriented Websites on Colorectal Cancer Screening, Clin Gastroenterol Hepatol. 2016 Jul 9 doi: 10.1016/j. cgh.2016.06.029. [Epub ahead of print].

14. Eline H. Schreuders, **Esmée J. Grobbee,** Manon C.W. Spaander, Ernst J. Kuipers. Advances in fecal tests for colorectal cancer creening. Current Treatment Options in Gastroenterology. 2016 Mar;14(1):152-62.

15. Eline H. Schreuders, **Esmée J. Grobbee**, Atija Kapidzic, Aafke H.C. van Roon, Anneke J. van Vuuren, Iris Lansdorp-Vogelaar, Wolfert Spijker, Kirsten Izelaar, Marco J. Bruno, Ernst J. Kuipers, Manon C.W. Spaander. Comparison of multiple rounds ONE versus TWO sample fecal immunochemical testing for colorectal cancer screening. Manuscript submitted.

16. Els Wieten, **Esmée J. Grobbee**, Paul Didden, Arjun D. Koch, Ernst J. Kuipers, Marco J. Bruno, Manon C.W. Spaander. Prevalence and treatment of T1 colorectal carcinoma in a FIT-based colorectal cancer screening program. Manuscript submitted.

17. **Esmée J. Grobbee,** Denise Altena. Princess Margaret Cancer Centre. In: Jeroen Kemperman, Jeroen Geelhoed, Jennifer op 't Hoog (red). Briljante Businessmodellen in de Zorg (pp. 324-337). 201 Amsterdam: Academic Service.

18. **Esmée J. Grobbee**, Eline H. Schreuders, Ernst J. Kuipers. Are there effective screening programs in Europe? From: Multidisciplinary management of rectal cancer 2nd edition, 2016 – in press.

19. **Esmée J. Grobbee,** Eline H. Schreuders, Manon C.W. Spaander, Ernst J. Kuipers. Colorectal cancer screening and diagnosis. From: Textbook of hepato-gastroenterology. Part II: gastroenterology, 2016 – in press.

PhD portfolio

Name PhD student:	E.J. (Esmée) Grobbee
PhD period:	April 2013 – November 2016
Erasmus MC department:	Gastroenterology and Hepatology
Promotor:	Prof. dr. Ernst J. Kuipers
Co-promotor:	Dr. Manon C.W. Spaander

Courses and workshops

	Year	Workload
EndNote workshop, Erasmus MC, Rotterdam	2013	6 hours
Pubmed workshop, Erasmus MC, Rotterdam	2013	6 hours
Reasearch Management for PhD students, Molecular medicine postgraduate school, Rotterdam	2013	28 hours
Cochrane Diagnostic Test Accuracy Review, authors' course Module 3, Cochrane Centre Birmingham, UK	2013	28 hours
Introduction to data-analysis, Erasmus Summer Programme, Rotterdam	2013	20 hours
Methods of public health research, Erasmus Summer Programme, Rotterdam	2013	20 hours
Practice of Epidemiologic Analysis, Erasmus Summer Programme, Rotterdam	2013	20 hours
Survival analysis course, Molecular medicine postgraduate school, Rotterdam	2013	12 hours
Integrity in science for PhD researchers, Dept of EMdical Ethics an philosophy, Erasmus MC, Rotterdam	2014	9 hours
Writing a Cochrane Diagnostic Test Accuracy Review, Julius UMC Utrecht	2014	16 hours
Biomedical English writing and communication, Erasmus MC, Rotterdam	2014	40 hours
Abdominal ultrasonography course, Dutch liver week 2012, Rotterdam	2012	10 hours
Indesign workshop, Molecular medicine postgraduate school, Rotterdam	2015	4 hours
Photoshop & Illustrator workshop, Molecular medicine postgraduate school, Rotterdam	2015	8 hours

Oral presentations at international conferences

	Year	Workload
Practice, indication and predictive factors of second look colonoscopy in a screening population. United European Gastroenterology Week, Vienna, Austria	2014	28 hours
Guaiac-based faecal occult blood tests versus faecal immunochemical test for colorectal cancer screening in average-risk individuals. World Endoscopy Organization (WEO) Colorectal Cancer Screening Meeting, Barcelona, Spain	2015	32 hours
Comparison of OC-sensor and FOB-Gold in population-based colorectal cancer screening based on FIT. Digestive Disease Week, Washington D.C., United States of America	2015	32 hours

CRC screening in the Netherlands. Problems and solutions. IV symposium of molecular pathology "Advances in the management of breast and colorectal cancer", Madrid, Spain	2016	32
Comparison of colonoscopy, sigmoidoscopy and multiple rounds of FIT-based colorectal cancer screening: long-term follow-up. World Endoscopy Organization (WEO) Colorectal Cancer Screening Meeting, San Diego, United States of America & Digestive Disease Week, San Diego, United States of America ***Certificate of recognition***	2016	40
Concentration of a Negative FIT Predicts Risk of Future Advanced Neoplasia: A Long-Term Follow-Up Study of Population-Based FIT Screenees. Digestive Disease Week, San Diego, United States of America ***Certificate of recognition***	2016	32

Oral presentations at national conferences

	Year	Workload
Practice, indication and predictive factors of second look colonoscopy in a screening population. Najaarscongres, Nederlandse Vereniging voor Gastro-enterologie, Velhoven, The Netherlands	2014	12 hours
FIT-based colorectal cancer screening: do we need to tailor screening for men and women? Voorjaarscongres, Nederlandse Vereniging voor Gastro-enterologie, Velhoven, The Netherlands	2015	12 hours
Comparison of OC-sensor and FOB-Gold in population-based colorectal cancer screening based on FIT. Voorjaarscongres, Nederlandse Vereniging voor Gastro-enterologie, Velhoven, The Netherlands	2015	12 hours
Comparison of colonoscopy, sigmoidoscopy and multiple rounds of FIT-based colorectal cancer screening: long-term follow-up. Voorjaarscongres, Nederlandse Vereniging voor Gastro-enterologie, Velhoven, The Netherlands	2016	12 hours
Concentration of a Negative FIT Predicts Risk of Future Advanced Neoplasia: A Long-Term Follow-Up Study of Population-Based FIT Screenees. Voorjaarscongres, Nederlandse Vereniging voor Gastro-enterologie, Velhoven, The Netherlands	2016	12 hours

Poster presentations at (inter)national conferences

	Year	Workload
Bowel preparation and hospital are important factors influencing quality of colonoscopy as measured by cecum intubation and adenoma detection. Digestive Disease Week, Chicago, United States of America	2014	12 hours
Fecal hemoglobin level is an important factor in predicting the risk of a second look colonoscopy in a screening population. Digestive Disease Week, Chicago, United States of America	2014	12 hours

FIT-based colorectal cancer screening: do we need to tailor screening for men and women? Digestive Disease Week, Washington D.C., United States of America ***Poster of excellence***	2015	12 hours
Guaiac-based faecal occult blood tests versus faecal immunochemical test for colorectal cancer screening in average-risk individuals. United European Gastroenterology Week, Barcelona, Spain	2015	12 hours

Attended (inter)national conferences

	Year	Workload
Najaarscongres, Nederlandse Vereniging voor Gastro-enterologie, Velhoven, The Netherlands	2013	12 hours
Digestive Disease Week, Chicago, United States of America	2014	28 hours
Najaarscongres, Nederlandse Vereniging voor Gastro-enterologie, Velhoven, The Netherlands	2014	12 hours
United European Gastroenterology Week, Vienna, Austria	2014	28 hours
Digestive Disease Week, Washington D.C., United States of America	2015	28 hours
United European Gastroenterology Week, Barcelona, Spain	2015	28 hours
Voorjaarscongres, Nederlandse Vereniging voor Gastro-enterologie, Velhoven, The Netherlands	2015	12 hours
Digestive Disease Week, San Diego, United States of America	2016	28 hours
Voorjaarscongres, Nederlandse Vereniging voor Gastro-enterologie, Velhoven, The Netherlands	2016	12 hours
United European Gastroenterology Week, Vienna, Austria	2016	28 hours

Attended seminars

	Year	Workload
3e Nationaal congres bevolkingsonderzoek darmkanker, Utrecht, the Netherlands	2013	6 hours
28th Erasmus Liver day. Rotterdam, The Netherlands	2013	6 hours
Wetenschapsmiddag Arts-assistenten vereniging (AAV), Erasmus MC, Rotterdam	2014	6 hours
World Endoscopy Organization (WEO) Colorectal Cancer Screening Meeting, Chicago, United States of America	2014	8 hours
6e Lagerhuisdebat Hepatitis B and C, Utrecht, the Netheralnds	2014	2 hours
World Endoscopy Organization (WEO) Colorectal Cancer Screening Meeting, Washington D.C., United States of America	2015	8 hours
World Endoscopy Organization (WEO) Colorectal Cancer Screening Meeting, Barcelona, Spain	2015	8 hours
11e Jaarsymposium Gastro-enterologie, Amsterdam	2016	6 hours
World Endoscopy Organization (WEO) Colorectal Cancer Screening Meeting, San Diego, United States of America	2016	8 hours

Awards

	Year
United European gastroenterology week young investigator bursary	2014
United European gastroenterology week young investigator bursary	2015
United European gastroenterology week young investigator bursary	2016

Memberships

	Year
Netherlands Association of Gastroenterology (NVGE)	2013-present
American Gastroenterological Assocation (AGA)	2013-2014

Review activities

Netherlands journal of medicine, BMC gastroenterology, Trials, Journal of Medical Screening

Educational activities and lecturing

	Year	Workload
Regioreferaat AAV, This is the end – het levenstestament (member of organizational committee and chair), Erasmus MC, Rotterdam	2015	28 hours
Wetenschapsmiddag Arts-assistenten vereniging (AAV; member of organizational committee), Erasmus MC, Rotterdam	2015	40 hours
Booklet practical guidelines; data management and quality for PhD-researchers	2016	12 hours


FOKKE & SUKKE
DOEN MEE AAN DE PILOT
WOW!
DEZE MOET JE OPSTUREN AAN HET DARMKANKER-ONDERZOEK!
RGvT

Dankwoord

Met deze laatste pagina's van dit boekje sluit ik een mooie periode af. Drie jaren waarin ik ontzettend veel geleerd heb, bijzondere en inspirerende mensen heb mogen ontmoeten en drie jaren waarop ik hoop verder te mogen bouwen. Want dit proefschrift vormt slechts een stap in een opleiding die nooit volledig zal zijn. Ik had dit niet kunnen doen zonder de steun, het vertrouwen en de hulp van vele mensen en een aantal van hen wil ik graag in het bijzonder bedanken.

Hooggeleerde prof. dr. Ernst J. Kuipers, beste Ernst, wat een geluk om onder jouw hoede te mogen promoveren. Als geen ander weet jij te inspireren en te motiveren. Ik kijk terug op vele mooie overlegmomenten, waar ik altijd nog enthousiaster (hoewel ook vaak met wat meer werk) vandaan kwam. Bij deze overleggen voelde ik me vrij om de discussie met je aan te gaan en wist jij me vaak de ogen te openen bij lastige problemen waarvoor jij (ogenschijnlijk) eenvoudige oplossingen wist te vinden. Veel dank voor je steun en begeleiding van dit proefschrift en ik hoop in de toekomst nog veel met je van gedachte te mogen blijven wisselen.

Lieve Manon (dr. Manon C. W. Spaander), na mijn eerste jaar promoveren werd jij mijn co-promotor, maar voor mijn gevoel was jij er al vanaf het begin. Gelukkig heb jij je door Ernst laten overtuigen om het CRC-team (i.e., het E-team) te begeleiden! Jouw kracht om van 'gewone' manuscripten, sterke stukken te maken heeft mij veel geleerd. En hoe jij de wetenschap (met de bijbehordende concurrentie) met de dagelijkse kliniek als dokter met geduld en vriendelijkheid kan combineren vind ik inspirerend. In de afgelopen jaren waren er naast gezellige en leerzame overlegmomenten, ook mooie avonden tijdens etentjes of borrels op congressen waar we ook over zaken buiten het ziekenhuis konden praten. Veel dank en ik zie er naar uit om over een paar jaar in het Erasmus MC ook met jou in de kliniek te mogen werken.

Een belangrijk onderdeel van dit onderzoek vormde de CRC-stuurgroepvergaderingen waarbij verschillende instanties, specialisaties en bijbehorende persoonlijkheden elkaar aanvulden om zo het maximale eruit te halen. Beste Iris (dr. I. Lansdorp-Vogelaar), veel dank voor je scherpe inzichten, het prettige overleg en je talent om de methodologie binnen een manuscript altijd naar een hoger niveau te tillen. Dank dat je plaats wilt nemen in de commissie. Beste Hanneke (dr. A.J. van Vuuren), veel dank voor je inzet in het laboratoriumonderzoek en je aanvullingen en inzichten tijdens stuurgroepvergaderingen. Beste Wolfert Spijker, Kristen Izelaar, Hans 't Mannetje en Jopie Krimpen, zonder jullie bestaat er geen proefbevolkingsonderzoek, heel veel dank voor jullie hulp

en ondersteuning. Beste Kirsten, jij bent pragmatisch en doortastend. Tijdens vergaderingen zag jij vaak de praktische oplossingen en wist jij als geen ander de hoofd- en bijzaken onder de aandacht te brengen, dit gold voor zowel de actuele problemen als het voorzien van toekomstige issues. Beste Hans, in momenten van chaos en paniek wist jij altijd de kalmte te bewaren en kon je precies die gegevens leveren waar ik naar op zoek was. Ook bij mijn vele hulpverzoeken over aantallen van testen, of vermiste deelnemers bleef jij altijd vriendelijk en heb je me heel veel geholpen, dankjewel!

Beste Agnes, veel dank voor je hulp met de patiënten op de polikliniek en voor je hulp rondom de logistiek van het proefbevolkingsonderzoek in het ziekenhuis. Uiteraard wil ik ook de endoscopie-verpleegkundigen, poli-assistenten, in het bijzonder Ronald, en endoscopisten, in het bijzonder Paul Didden, bedanken.

Dank past ook aan alle deelnemers van het proefbevolkingsonderzoek. Zonder jullie inzet waren er geen resultaten geweest, waarbij genoemd moet worden dat op ieder congres er met ontzag wordt gesproken over de ongelooflijk hoge deelnamegraad in de studie.

I would also like to thank all (international) co-authors for the positive and fruitfull collaborations.

Graag wil ik prof. dr. M.J. Bruno, prof dr. P.M. M. Bossuyt en prof dr. J.F. Lange hartelijk danken voor het beoordelen van mijn proefschrift en het plaastnemen in de kleine commissie. Prof. dr. Evelien Dekker, dank voor het zitting nemen in de promotiecommissie, ik bewonder jouw kracht en gedrevenheid en heb veel van je geleerd in de afgelopen jaren. Prof. dr. Stephan P. Halloran. I am hounoured that you are willing to take place in the PhD committee for my thesis defence.

Beste prof dr. R.A. de Man, veel dank voor het in mij gestelde vertrouwen door mij op te leiden tot Maag-, Darm-, en Leverarts.

Best dr. H. Boom, dr J. T. Brouwer en alle collega arts-assistenten en medisch specialisten van het Reinier de Graaf Gasthuis. Veel dank voor het warme welkom en het fijne opleidingsklimaat, ik kijk uit naar de komende drie jaar als arts-assistent in dit ziekenhuis.

Lieve Manon, ook wel de Amsterdamse Esmée genoemd, of was ik nou juist de Rotterdamse Manon? Samen gingen we dit onderzoek aan en hebben we over de jaren veel overleg gehad en bijzondere momenten meegemaakt. Jouw klinische blik en mijn voorliefde voor SPSS maakte ons een goede

combinatie. Gelukkig mogen we de komende tijd nog blijven samenwerken en blijven we ook in de toekomst MDL-collega's. Ik heb nu al zin in de dag van jouw promotie.

Er is mogelijk geen inspirerender omgeving dan het DAK. Lieve dak-duifjes & lab-ratjes, julllie hebben mijn promotietijd onvergetelijk gemaakt. Vele mooie momenten hebben we gedeeld, van zwetend zwoegen op het DAK tijdens zinderende zomers, tot borrels in de stad en de trips en feestjes tijdens vele congressen. Dankjulliewel!

Lieve Eline, jij was degene die me wegwijs maakte in de CRC-screeningswereld en na onze trip naar Birmingham wist ik dat ik naast een hele leuke collega, er ook een vriendin bij had. Juist omdat we zo verschillend zijn heb ik heel veel van je geleerd. Dankjewel voor je vriendschap en geduld. Ik weet zeker dat jij binnekort zal schitteren tijdens jouw verdediging en dat je een fantastische huisarts zult worden.

Lieve Els en Sil, ik voel me een geluksvolgel met jullie twee vandaag aan mijn zijde. Veel dank voor jullie steun, jullie eerlijkheid en alle gezelligheid. Ik kijk uit naar onze toekomst als MDL-collega's, en weet zeker dat gedurende onze carrière (en wie weet zelfs daarna) we nog heel veel mooie momenten zullen delen, welke verder reiken dan alleen ons werk. Dankjulliewel dat jullie mijn paranifm willen zijn op deze dag.

Lieve JB, dankjewel dat je mijn wereld verbreedt. Avec toi, on est vraiment mieux que d'être en prison.

Lieve broertjes, ik ben onzettend trots om jullie zus te mogen zijn. Lieve Thom, dank voor je steun en humor, ik heb bewonderding voor je talent om mensen te laten groeien en leren, ik weet zeker dat jij een inspiratie zal zijn voor vele mensenlevens. Lieve Daan, dank voor je creatieve bijdrage aan de lay-out van dit boekje. Ik bewonder hoe jij je eigen pad kiest, met jouw creativiteit en talent kom je er zeker! Lieve pap en mam, jullie liefde voor ons, maar vooral julie liefde voor elkaar is het mooiste cadeau dat jullie mij ooit konden geven. Lieve mam, jij houdt me met beide benen op de grond en laat me zien hoe veel moois het leven te bieden heeft. Lieve pap, het is bijzonder dat ik 30 jaar later aan de zelfde universiteit mijn proefschrift mag verdedigen en voel me trots dat ik hier vandaag als jouw dochter mag staan. Veel dank voor jullie onvoorwaardelijke steun en voor jullie bijdrage welke zich niet laat samenvatten in deze laatste zinnen.

About the author

Elisabeth Jacomina (Esmée) Grobbee was born on August 19th 1987 in Nieuwegein, the Netherlands. She attended the Gymnasium at the Christelijk Lyceum in Zeist and graduated in 2005. In that same year she started at the Roosevelt Academy in Middelburg (direction life sciences). The following year she started medical school at the Leiden University Medical Centre. As part of her studies she took part in research projects abroad at the University Hospital in Kuala Lumpur, Malaysia and at the Monash University in Melbourne, Australia. During the last part of her medical training, the rotations, she spent time in Paramaribo at a general practitioner office , Suriname. In 2013 she obtained her medical degree and started her PhD trajectory at the Erasmus MC University Medical Center in Rotterdam at the department of Gastroenterology and Hepatology under the supervision of prof. dr. E.J. Kuipers and dr. M.C.W. Spaander. As of July 2016 she has started her two-year Internal Medicine residency (program director dr. H. Boom) and following two-year Gastroenterology and Hepatology residency (program director dr. J.T. Brouwer) at the Reinier de Graaf hospital. Hereafter, as part of the formal postgraduate training in Gastroenterology and Hepatology (cluster Erasmus MC University Medical Centre, Rotterdam, program director prof. dr. R.A. de Man) she will continue with the last two years of her training at the Erasmus MC University Medical Centre.

from FIT to Future - 2016