Optimizing colorectal cancer screening using fecal immunochemical tests

de Klerk, C.M.

Publication date
2019
Document Version
Other version
License
Other

1

EQUIVALENT ACCURACY OF 2 QUANTITATIVE FECAL IMMUNOCHEMICAL TESTS IN DETECTING ADVANCED NEOPLASIA IN AN ORGANIZED COLORECTAL CANCER SCREENING PROGRAM

C.M. de Klerk*
E. Wieten*
A. van der Steen
C.R. Ramakers
E.J. Kuipers
B.E. Hansen
I. Lansdorp-Vogelaar
P.M. Bossuyt
E. Dekker
M.C.W. Spaander

*shared first author

Gastroenterology 2018

1

ABSTRACT

Background

Although different brands of fecal immunochemical tests (FITs) are used for colorectal cancer (CRC) screening, few studies have compared their accuracy in detecting advanced neoplasia.

Methods

We performed a large prospective cohort study within the Dutch national CRC screening program to evaluate 2 quantitative FITs: FOB-Gold (Sentinel, Italy) and OC-Sensor (Eiken, Japan, reference standard), from May 2016 through March 2017. We randomly selected 42,179 screening-naïve individuals (55–75 years old), who were asked to perform both FITs themselves using the same bowel movement. Participants with positive results from 1 or both FITs (≥15 µg hemoglobin/gram feces) were invited for colonoscopy examination. Equivalence in detection of advanced neoplasia was evaluated with a predefined margin of 0.15%.

Results

Of 42,179 invitees, 22,064 (52%) participated and FITs were completed for 21,078 participants. Of 2112 participants (9.6%) with 1 or 2 positive results from FITs, 1778 (84%) underwent a colonoscopy. Of all invitees, the FOB-Gold test detected advanced neoplasia (confirmed by colonoscopy) in 610 participants (1.45%) and the OC-Sensor detected advanced neoplasia (confirmed by colonoscopy) in 606 participants (1.44%), an absolute difference of 0.01% (95% confidence interval [CI], −0.06% to 0.08%). Of the 21,078 participants who completed both FITs, 1582 (7.5%) had a positive result from the FOB-Gold test and 1627 (7.7%) a positive result from the OC-Sensor test (P=.140). The relative true-positive rate of FOB-Gold vs OC-Sensor in detecting advanced neoplasia was 0.97 (95% CI, 0.92–1.01) and 0.95 (95% CI, 0.87–1.03) for CRC. The relative false-positive rate of the FOB-Gold test vs the OC-Sensor test in detecting advanced neoplasia was 0.99 (95% CI, 0.93–1.05)

CONCLUSION

In a large prospective study of individuals invited for CRC screening in The Netherlands, we found equivalent accuracy of the FOB-Gold FIT vs the OC-Sensor FIT in detecting advanced neoplasia. These results are relevant for selecting FITs for CRC screening programs worldwide. Dutch National Trial Registry: NTR5874

INTRODUCTION

Colorectal cancer (CRC) is a significant public health problem due to its high incidence, morbidity and mortality.[1] It is the third most commonly diagnosed cancer in men and second in women.[1] CRC screening by means of guaiac fecal occult blood testing (gFOBT) and sigmoidoscopy has been shown to reduce CRC-related mortality through the detection of advanced adenomas and early-stage CRC.[2-6] Accordingly, CRC screening programs are being implemented worldwide.[7]

Fecal immunochemical tests (FITs) are rapidly replacing gFOBTs for CRC screening, because of their easier handling, higher uptake, automated assessment, and higher sensitivity in detecting advanced neoplasia (AN).[8] In addition, FITs can provide a quantitative test result, allowing cutoff adjustment to match available colonoscopy resources. FITs also allow for single stool testing, and do not require dietary restrictions, and, consequently, lead to higher participation rates than gFOBTs.[8-11] As a consequence, many countries are either in the transition from gFOBT to FIT screening, or implement FIT screening from the start.[7]

In the European guidelines, the quantitative immunochemical tests are recommended as test of choice for population screening.[9] These guidelines also indicate that a screening program should assess individual device characteristics, including accuracy, ease of use by participant and laboratory, suitability for transport, sampling reproducibility and sample stability. However, comparative data with enough power from head-to-head comparisons to decide on equivalence of FIT assays in detecting AN in population screening are not yet available, hindering informed decision making.

At present, multiple FITs are available, of which FOB-Gold (Sentinel, Milan, Italy) and OC-Sensor (Eiken Chemical, Tokyo, Japan) are currently widely used.[12] FIT assays vary in analytical performance due to a range of factors, including anti-heme antibody characteristics and assay optimization, buffer composition and volume and sample tube design. These differences influence the measured fecal hemoglobin (Hb) concentration, FIT positivity rate, error rates, and capacity to detect AN.[13, 14] One randomized controlled trial reported higher true-positivity and false-positivity rates for FOB-Gold compared to OC-Sensor at equal positivity cutoffs (in ng Hb/mL buffer).[15] Three other randomized trials reported equal AN detection rates of both tests, but higher positivity rates with FOB-Gold.[13, 16, 17] However, all these studies were small and randomized invitees to one of both tests.[13, 15, 16] A recent comparative evaluation of 9 FITs, including OC-Sensor and FOB-Gold, showed that sensitivity to detect AN varied widely between FITs, ranging from 16% to 34%.[18] Yet, this retrospective study was performed in a laboratory setting using stored samples and limited inclusions per test. A large population-based study with a paired design, in which both tests are compared within the same individual and sampled from the same stool would minimize the risk of confounding factors and increase the applicability of the study results to CRC screening programs.

We therefore conducted a large prospective population-based study within the Dutch nationwide CRC screening program to compare the accuracy in detecting AN for 2 widely used FITs.

1

METHODS

Study Population

This study was conducted within the Dutch CRC screening program between May 2016 and March 2017. The structure of the Dutch CRC screening program has been described elsewhere.[19] In short, demographic data of all individuals between the ages of 55 and 75 years, living in the southwest region of the Netherlands, were obtained from municipal registers. The target population for our study consisted of first-time invitees in 2016, that is those aged 59, 61, 63, 71, or 75 years.[19] This population encompassed more than 250,000 eligible screening-naïve individuals. For the purpose of this study, a random sample was taken from this target population with a computer run algorithm (SPSS, version 23, IBM Corp, Armonk, NY). Selection of study invitees preceded invitation. This study was ethically approved by the Minister of Health (Population Screening Act; no. 769500-1357 16-PG).

Study design and intervention

Two FITs (one FOB-Gold and one OC-Sensor) were sent by postal mail. Individuals were invited to collect a single feces sample of the same bowel movement with each test. Detailed sampling instructions as recommended by the manufacturer were provided for each test. Study invitees were asked to fill out an informed consent form, including the sampling date. After 42 days, a reminder was sent automatically to nonresponders. Consenting invitees were asked to return both tests and consent form in a prepaid and sealed envelope to one accredited, centralized laboratory. Persons who had actively deregistered from the screening program, who had moved, were deceased, did not consent or did not respond to the study invitation were labelled as nonparticipants.[19]

Appendix 1a and 1b describe the pre-analytical aspects and analytical performance of the FIT analysis. The OC-Sensor tests were analyzed by using the OC-Sensor Diana analyzer. The FOB-Gold tests were analyzed by using the Bio Majesty JCA-BM6010/C analyzer. Quantitative results for both tests were provided in ng Hb/mL. For the purpose of the study, FOB-Gold test were considered positive at ≥88 ng Hb/mL and OC-Sensor tests at ≥75 ng Hb/mL. At the time that the study was designed, a positivity cut-off of ≥88 ng Hb/mL was used in the Dutch nationwide CRC screening program for FOB-Gold tests. To be able to compare FOB-Gold to OC-Sensor tests within the nationwide CRC screening program, OC-Sensor test positivity cut-off (in µg Hb/g feces) was set to be the same as that for FOB-Gold. Converted into micrograms (µg) of Hb per gram of stool, the threshold for a positive test result was ≥15 µg Hb/g feces for both tests.[20]

Participants were referred for colonoscopy in case of 1 or 2 positive FIT results. Participants with 2 negative test results were referred back to the Dutch CRC screening program and will be re-invited after 2 years. Participants who returned 2 non-analyzable tests, who had 2 unreliable test results, or for whom both tests were missing, were referred back to the Dutch CRC screening program and were not included for the primary outcome of this study.

Participants were informed about their FIT results by postal mail within 5 days after the FIT was analyzed. If 1 or both FIT results were positive, the family physician was informed and the participant

was invited for a precolonoscopy interview in an accredited colonoscopy center nearby the participant's home address. At this precolonoscopy interview, participants' eligibility for colonoscopy was assessed. Colonoscopy exclusion criteria were similar to those of the Dutch CRC screening program: a life expectancy of 5 years or less, a proctocolectomy in the past, under current treatment for CRC, history of inflammatory bowel disease, and a complete colonoscopy in the past 5 years.[21] Colonoscopies were performed within 10 days after the interview at one of the certified colonoscopy centers by accredited endoscopists who performed at least 200 colonoscopies a year with an adenoma detection rate of ≥ 30%.

Location, size and morphology of all identified colorectal lesions were reported using an automated structured colonoscopy reporting system. Polyps were removed and sent for pathology review in separate jars.[22] Advanced adenoma was defined as an adenoma ≥ 10 mm, with ≥ 25% villous component and/or high-grade dysplasia. AN included advanced adenoma or CRC. If multiple lesions were present, the participant was classified according to the most advanced lesion. Logistics were executed conform the Dutch CRC screening quality guidelines.[19 23] Socioeconomic status was assessed using the Dutch area social status score, grouped into quintiles. These scores are developed by the Netherlands Institute of Social Research, and are a composite measure of education, income and employment.[24]

Outcome measures and statistical analysis

Our primary outcome measure was the difference in diagnostic yield of AN between OC-Sensor and FOB-Gold, defined as the number of participants with AN detected relative to the number of invitees.

We hypothesized that the 2 FITs would generate an equivalent diagnostic yield, defined as an absolute symmetric difference of at most 0.15%, relative to the number of invitees. Differences in paired proportions of diagnostic yield were evaluated using the method described by Liu et al.[25] Confidence intervals (CI) were calculated using a Wald interval with Bonett-Price adjustment.[26] Comparisons in diagnostic yield for CRC between both tests were evaluated similarly.

In addition, we compared the accuracy of both FITs in participants with paired test results, that is, in participants with 2 complete and reliable FIT results. Excluded from paired analysis were participants with 1 or 2 non-analyzable tests (due to fecal overload, loss of two-thirds or more of the total buffer volume both from visual assess by laboratory staff and by automatic system, missing barcode or another technical problem) and participants with an unreliable test result (in case the return date was more than 6 days after sampling or if the sampling date was missing). Paired accuracy was assessed by calculating the relative true-positive rate and the relative false-positive rate. The relative true-positive rate is defined as the number of true-positive results (in whom AN was detected at colonoscopy) for FOB-Gold relative to the number of true-positive results for OC-Sensor. The relative true-positive rate is equal to the relative sensitivity rate in screening participants. A relative sensitivity of 1.00 implies that both screening tests result in the same sensitivity, a relative

sensitivity of 1.10 would imply that using FOB-Gold results in 10% more true-positives. The relative true-positive rate was calculated with 95% CIs using the methods proposed by Alonzo et al.[27] Similarly, the relative false-positive rate was estimated, defined as the number of false-positive test results (in whom no AN was detected at colonoscopy) for FOB-Gold relative to the number of false-positive test results for OC-Sensor.[27]

Our secondary outcomes included the number of analyzable tests, the positivity rate (defined as the proportion of participants with a positive FIT result) and the positive predictive values (defined as the number of participants in whom AN was detected relative to those undergoing colonoscopy after a positive FIT result). The participation rate was calculated as the number of participants relative to the number of invitees.

Paired proportions were compared using the McNemar test. Other proportions were compared using chi-square statistics. All p values were 2-sided; differences were considered significant if $p<0.05$.

Sample size calculations

In this equivalence trial, we assumed a 1.5% diagnostic yield for AN with both tests, anticipated no difference, and wanted to exclude an absolute difference in diagnostic yield of 0.15% or more between the 2 tests. We expected that 5 per 1000 invitees would have discordant true-positive FIT results: invitees with AN detected after 1 (not 2) positive FIT results. In that case, inviting 40,000 screenees, with an expected participation rate of 50% (20,000 screenees with 2 complete test returns), would give us more than 90% power to demonstrate equivalence, using an alpha of 0.05.[25] All authors had access to the study data and reviewed and approved the final manuscript.

RESULTS

Cohort characteristics

Of 42,179 individuals invited for screening, 22,064 (52%) participated in the study (Figure 1). Baseline characteristics of study participants are shown in Table 1. Of the participants, 21,078 (96%) completed both FITs in a single bowel movement, 694 (3.2%) completed only one FIT, and 292 returned two incomplete tests (non-analyzable tests or tests with unreliable results).

All 2112 (9.6%) with 1 or 2 positive FIT results were invited for a precolonoscopy interview; 1778 (84%) underwent a colonoscopy. Of those who attended colonoscopy 1066 were male (60%) and median age was 60 years (interquartile range 59-63 years; Table 1). In 716 (4.0%) participants with a positive FIT result who underwent a colonoscopy AN was detected and in 82 (0.5%) CRC was detected.

Diagnostic yield for advanced neoplasia and colorectal cancer

AN was detected after a positive FOB-Gold in 610 of total invitees (1.45%) and in 606 (1.44%) after a positive OC-Sensor, resulting in an absolute difference of 0.01% (95%CI -0.06% to 0.08%).

Figure 1. Study flow

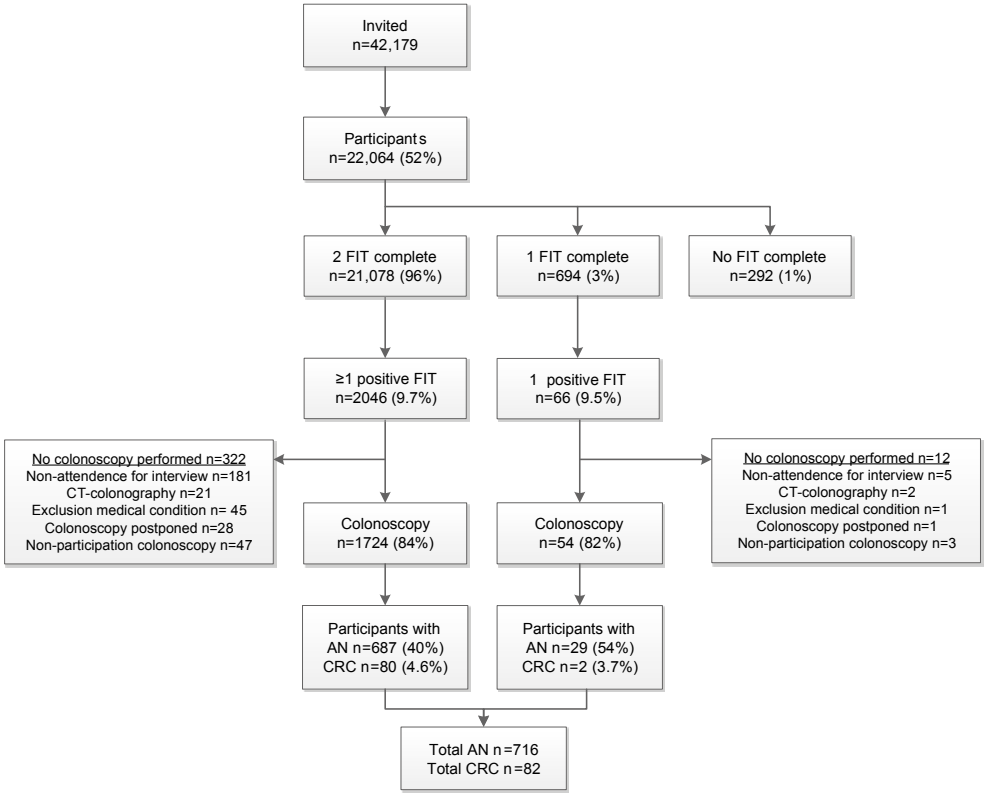


Table 1. Characteristics of study participants at baseline, with ≥1 positive fecal immunochemical test results, and at colonoscopy

Total	Participants n=22,057	Participants with ≥1 positive FIT result n=2112	Participants with colonoscopy n=1778
Male sex, n (%)	10,589 (50)	1256 (60)	1066 (60)
Age*, median (IQR)	60 (58-62)	60 (59-63)	60 (59-63)
Socioeconomic status, n (%)			
Very high	4407 (20)	397 (19)	341 (19)
High	4843 (22)	387 (18)	330 (19)
Average	4052 (18)	372 (18)	316 (18)
Low	5136 (23)	536 (25)	451(25)
Very low	3577 (16)	416 (20)	336 (19)
Missing	49 (<1)	4 (<1)	4 (<1)

*Age at time of FIT invitation
IQR=interquartile range
Socioeconomic status scores are a composite area-based measure of education, income and employment.

The existence of a difference of 0.15% or more between FOB-Gold and OC-Sensor in detection of AN was rejected (p < 0.001). CRC was detected after a positive FOB-Gold result in 74 invitees (0.18%) and after a positive OC-Sensor result in 78 (0.18%), resulting in an absolute difference of -0.009% (95%CI -0.027% to 0.008%).

Incomplete test returns and non-analyzable tests

Differences between FOB-Gold and OC-Sensor tubes in the proportion of non-analyzable results are shown in Table 2. Forty-nine of the 22,057 (0.22%) returned FOB-Gold tests were non-analyzable vs 14 of 21,369 returned OC-Sensor tests (0.07%; p<0.001). Buffer loss was more frequently a problem for analysis of FOB-Gold tests than for OC-Sensor (p=0.005).

Paired accuracy: relative true-positive rate and relative false-positive rate

The results of 986 (4.5%) participants were not included in the paired analysis due to not returning 1 of the tests (n=702), ≥1 non-analyzable test(s) (n=54) or ≥1 unreliable test result(s) (n=260), or a combination of these (n=30). 695 participants did not provide a OC-Sensor test result and 7 an FOB-Gold test result. Of the 21,078 participants who completed both FITs, the positivity rate was 7.5% for FOB-Gold compared to 7.7% for OC-sensor (p=0.140) (Table 3). A total of 1163 (57%) had 2 positive FITs, 419 (20%) participants had a positive FOB-Gold and negative OC-Sensor result, and 464 (23%) participants a negative FOB-Gold and positive OC-Sensor result.

The relative true-positive rate (relative sensitivity) for the detection of AN for FOB-Gold relative to OC-Sensor was 0.97 (95%CI: 0.92 to 1.01; Figure 2), and the false-positive rate was 0.99 (95%CI: 0.93 to 1.05). Positive predictive value of AN was 43.8% for FOB-Gold and 44.3% for OC-sensor (absolute difference 0.5%; 95%CI: -3.3% to 4.2%, p <0.05). The relative true-positive rate (relative sensitivity) for FOB-Gold versus OC-Sensor in detecting CRC was 0.95 (95%CI: 0.87 to 1.03; Figure 2). Positive predictive value of CRC was 5.5% for FOB-Gold and 5.7% for OC-sensor (absolute difference 0.2%; 95%CI: -1.5% to 2.0%).

Table 2. Reasons for non-analyzable fecal immunochemical test tubes

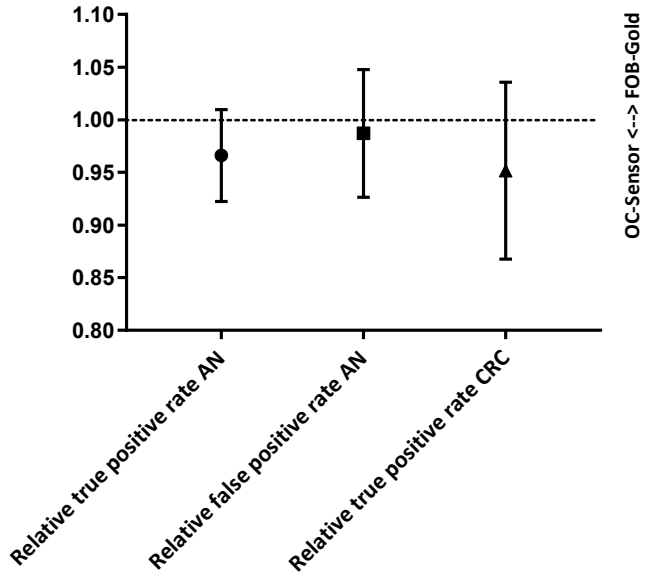
Total returned tests	**FOB-Gold n=22,057**	**OC-Sensor n=21,369**	**P-value**
Non-analysable tests, n (%)	**49 (0.22)**	**14 (0.07)**	**<0.001**
Technically impossible	10 (20)	4 (29)	0.52
Barcode unreadable	5 (10)	4 (29)	0.09
Broken tube	2 (4.1)	0	0.45
No buffer	23 (47)	1 (7..1)	0.005
Too large sample	6 (12)	1 (7.1)	0.60
Too small sample	3 (6.1)	3 (21)	0.09
No sample taken	0	1 (7.1)	0.06

Table 3. Detection of advanced neoplasia and colorectal cancer with FOB-Gold and OC-Sensor in paired design

		FOB-Gold		
		Positive*	**Negative**	**Total**
OC-Sensor	**Positive***	1163	464	1627 (7.7%)
	AN	500	104	604 (2.9%)
	CRC	70	7	77 (0.4%)
	Negative	419	19,032	19,451 (92.3%)
	AN	83	-	-
	CRC	3	-	-
	Total	1582 (7.5%)	19,496 (92.5%)	21,078
	AN	583 (2.8%)	-	-
	CRC	73 (0.4%)	-	-

Results for participants with two completed FITs only
AN: advanced neoplasia (CRC and/or advanced adenoma, defined as an adenoma ≥10 mm, with ≥25% villous component and/or high-grade dysplasia).
CRC: colorectal cancer
*Overall colonoscopy attendance rate of participants with ≥1 positive FIT(s) was 84% at fecal haemoglobin (Hb) positivity cut-off of ≥15 µg Hb/g feces.

Figure 2. Relative true-positive and false-positive rates FOB-Gold vs OC-Sensor with 95% CIs



1

DISCUSSION

This large prospective trial within the Dutch national CRC screening program shows that 2 widely used FITs, FOB-Gold and OC-Sensor, have similar accuracy in detecting AN and CRC at a positivity cut-off level of ≥15 µg Hb/g feces.

Factors adding to the validity of our results include the study's paired design, its implementation within the logistics of the nationwide screening program, and its large and representative cohort of participants. Moreover, both FITs were analyzed in the same laboratory at the day of arrival and followed identical logistic routes. Consequently, factors that are known to influence test results were identical, including temperature changes, time differences from sampling to analyzing, and laboratorial logistic differences.[28] Screening logistics, from invitation to pathology for those who underwent colonoscopy, were identical to those of the Dutch nationwide screening program.

Some potential limitations also have to be acknowledged. Participation rate in this study was lower (52%) than in the current Dutch nationwide screening program (72%) and previously performed pilot studies in our country (60-63%).[13, 29,] We had anticipated this difference since an alternative existed for our study invitees: those who opted out of the study could still participate in the regular one-FIT population CRC screening program. For this reason, we invited over 40,000 persons for the study, to have enough power and precision for a comparison of the 2 FITs. As no colonoscopies were performed in participants with 2 negative test results, we are unable to provide estimates of sensitivity, yet we calculated estimates of relative sensitivity. We also observed that more OC-Sensor tubes were missing than FOB-Gold tubes. This was most likely due to the way the study was embedded within the nationwide screening program. Study invitees were asked to perform both tests if they wished to participate in the study. Of those invitees that did not want to participate in the study but had a preference to participate in the regular nationwide screening program with 1 test, only the FOB-Gold was analyzed in the laboratory. The form was checked twice, first manually by laboratory assistants and subsequently by an automated system. In case no study approval was given by the screenee, only the FOB-Gold was analyzed by the laboratory assistant as this is the test that is currently used in the nationwide CRC screening program. However, this means some OC-Sensor tests may have been falsely qualified as being not returned by the screenee, while in reality the test was sent back with written study consent. This could have occurred when the manual check for informed consent by the laboratory staff differed from the scanned consent by the computer.

Previous studies comparing FIT brands in CRC screening studied either randomized subjects to perform a single FIT [13, 15-17], or had FITs performed in different bowel movements.[30] Differences in true-positive rate in favor of OC-Sensor to FOB-Gold were found in a Spanish screening cohort, in which participants were randomly invited to perform one of both tests.[15] The cut-off was equalized in ng Hb/mL buffer, instead of µg Hb/g feces, therefore subject to known differences in buffer volumes between both FIT brands (1.7 mL FOB-Gold and 2.0 mL OC-Sensor).[20] Another randomized trial in France showed equal detection rates for both FITs but found a lower positive predictive value

for FOB-Gold than OC-Sensor.[17] In a Latvian screening cohort, no difference between FOB-Gold and OC-Sensor was found in number needed to scope to detect 1 participant with AN at a positivity cut-off of 15 µg Hb/g faeces.[16] In a Dutch screening pilot, diagnostic yield and positive predictive value were equal for both FITs at equal cutoffs, and equal positivity rates.[13] The authors recommended to compare FIT brands at equal positivity rates, rather than equal cutoffs. In our study, the same cutoff yielded comparable positivity rates. In contrast to our study, these previous studies were not powered to determine small differences in the detection of AN.

Almost half of the participants with a positive test result had only 1 positive FIT result. One could wonder whether such discordant results will also be detected when 2 FITs of the same brand are compared. Discordant detection and positivity rates between 2 identical FITs might result from an uneven distribution of hemoglobin through the feces. It should also be noted that the sampling instructions by the manufacturers of FOB-Gold and OC-Sensor differ, which may influence test handling and, possibly, its results. FOB-Gold instructions prescribe inserting the sampling stick in 4 different parts of the stool. The OC-Sensor on the other hand, should be scraped through the stool in 4 different parts. Both sets of instructions were provided as such to our study invitees. We found, however, no evidence that these sample methods resulted in different detection rates.

In line with existing evidence, we found more sampling errors with the FOB-Gold than OC-sensor, although proportions in our study were smaller (0.22% vs 0.07%) than in the Dutch pilot screening program (2.0% vs 0.7%, $p<0.001$)[13], the Spanish study (2.3% vs 0.2%)[15], and within a Latvian screening setting (4.4% vs 0.2%, $p<0.001$)[31]. This might be explained by the fact that FOB-Gold adjusted its cap to prevent buffer loss by opening the wrong side of the tube. Furthermore, in the Dutch pilot screening program, screenees were probably more used to the OC-Sensor test because this test was used in the first 3 rounds and FOB-Gold was only introduced in the fourth. We found, however, no evidence that these sample methods resulted in different detection rates.

CRC is the second most common cancer and cause of cancer-related death worldwide.[1] To reduce CRC incidence and mortality in Europe, the European Committee have made recommendations on CRC screening for all European Union members.[32] They stated that the desired level of screening coverage of the target population is 95% with a desirable level of uptake of >65%.To fulfil these target levels, one needs to have an organized population-based screening program with a high participation rate for the selected screening tool. Participation rates of FIT are generally high, therefore FIT has been selected in most countries as the screening test of choice.[7]

As we found no evidence that FOB-Gold and OC-Sensor differ in detecting AN and CRC, other features can now guide informed decision making when selecting 1 of these 2 brands for FIT-based CRC screening. Probably the next most important consideration is ease of use of the test and its effect on participation rate. As stated by Winawer and Allison: 'the best test is the one that gets done well'.[33] Many other factors should be considered including costs, ease of use for laboratory staff or other stakeholders involved in FIT analysis, suitability for transport, the keeping quality

of the tubes, analyzer features, capacity, speed, analytical performance, sample stability, easy of handling, safety during postage and labeling. Depending on context and setting, more studies are warranted to evaluate these other aspects of FIT.

When this trial started, the Dutch nationwide screening program was already launched. Yet evaluations of screening strategies should not be limited to the implementation phase. As others have argued, evaluation should continuously be explored within and while running a screening program, with formal study designs such as paired comparisons or randomized trials.[34] Apart from its results, our study may therefore serve as an example how to assess and possibly improve screening effectiveness within an ongoing program and may, hopefully, inspire future initiatives.

ACKNOWLEDGEMENTS

The authors would like to thank the Netherlands Organization for Health Research and Development of the Dutch Ministry of Health (ZonMw) for funding, and all participants for their efforts to perform two tests and a questionnaire. The authors thank all involved co-workers of the Foundation of Population Screening South-West (Bevolkingsonderzoek Zuid-West), and all involved co-workers at Daklapack packaging for their contributions to the study. Finally, the authors thank Star lab (Rotterdam, The Netherlands) for their efforts in analyzing twice as many FITs compared to the nationwide Dutch screening program during the study period.

REFERENCES

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 2015;136:E359-86.
2. Zauber AG, Winawer SJ, O'Brien MJ, Lansdorp-Vogelaar I, van Ballegooijen M, Hankey BF, et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. N Engl J Med 2012;366:687-96.
3. Hardcastle JD, Chamberlain JO, Robinson MH, Moss SM, Amar SS, Balfour TW, et al. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. Lancet 1996;348:1472-7.
4. 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:Cd001216.
5. 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;49:3049-54.
6. Holme O, Loberg M, Kalager M, Bretthauer M, Hernan MA, Aas E, et al. Effect of flexible sigmoidoscopy screening on colorectal cancer incidence and mortality: a randomized clinical trial. Jama 2014;312:606-15.
7. 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;64:1637-49.
8. Kuipers EJ, Spaander MC. Colorectal Cancer Screening by Colonoscopy, CT-Colonography, or Fecal Immunochemical Test. J Natl Cancer Inst 2016;108:djv383.
9. Halloran SP, Launoy G, Zappa M. European guidelines for quality assurance in colorectal cancer screening and diagnosis. First Edition--Faecal occult blood testing. Endoscopy 2012;44 Suppl 3:Se65-87.
10. Kuipers EJ, Rosch T, Bretthauer M. Colorectal cancer screening--optimizing current strategies and new directions. Nat Rev Clin Oncol 2013;10:130-42.
11. Digby J, McDonald PJ, Strachan JA, Libby G, Steele RJ, Fraser CG. Use of a faecal immunochemical test narrows current gaps in uptake for sex, age and deprivation in a bowel cancer screening programme. J Med Screen 2013;20:80-5.
12. Schreuders EH, Grobbee EJ, Spaander MCW, Kuipers EJ. Advances in Fecal Tests for Colorectal Cancer Screening. Current Treatment Options in Gastroenterology 2016;14:152-62.
13. Grobbee EJ, van der Vlugt M, van Vuuren AJ, Stroobants AK, Mundt MW, Spijker WJ, et al. A randomised comparison of two faecal immunochemical tests in population-based colorectal cancer screening. Gut 2016;66:1975-1982.
14. Chiang TH, Chuang SL, Chen SL, Chiu HM, Yen AM, Chiu SY, et al. Difference in performance of fecal immunochemical tests with the same hemoglobin cutoff concentration in a nationwide colorectal cancer screening program. Gastroenterology 2014;147:1317-26.
15. Zubero MB, Arana-Arri E, Pijoan JI, Portillo I, Idigoras I, Lopez-Urrutia A, et al. Population-based colorectal cancer screening: comparison of two fecal occult blood test. Front Pharmacol 2014;4:175.
16. Santare D, Kojalo I, Liepniece-Karele I, Kikuste I, Tolmanis I, Polaka I, et al. Comparison of the yield from two faecal immunochemical tests at identical cutoff concentrations - a randomized trial in Latvia. Eur J Gastroenterol Hepatol 2016;28:904-10.
17. 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. Eur J Cancer 2012;48:2969-76.
18. Gies A, Cuk K, Schrotz-King P, Brenner H. Direct Comparison of Diagnostic Performance of 9 Quantitative Fecal Immunochemical Tests for Colorectal Cancer Screening. Gastroenterology 2018;154:93-104.
19. Toes-Zoutendijk E, van Leerdam ME, Dekker E, van Hees F, Penning C, Nagtegaal I, et al. Real-Time Monitoring of Results During First Year of Dutch Colorectal Cancer Screening Program and Optimization by Altering Fecal Immunochemical Test Cut-off Levels. Gastroenterology 2016.

20. Fraser CG, Allison JE, Halloran SP, Young GP, Org WE. A Proposal to Standardize Reporting Units for Fecal Immunochemical Tests for Hemoglobin. Jnci-Journal of the National Cancer Institute 2012;104:810-14.
21. National Institute for Public Health and the Environment (RIVM). Exclusion criteria colonoscopy (Dutch) http://www.rivm.nl/Documenten_en_publicaties/Professioneel_Praktisch/Richtlijnen/Preventie_Ziekte_Zorg/Darmkanker/Exclusiecriteria_coloscopie (date accessed 20th June 2017), 2013.
22. Casparie M, Tiebosch AT, Burger G, Blauwgeers H, van de Pol A, van Krieken JH, et al. Pathology databanking and biobanking in The Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive. Cell Oncol 2007;29:19-24.
23. National Institute for Public Health and the Environment (RIVM). Uitvoeringskader Bevolkingsonderzoek Darmkanker (Dutch): www.bevolkingsonderzoekdarmkanker.nl/downloads (date accessed 20th June 2017), 2017.
24. Netherlands Institute of Social Research. Socioeconomic status scores The Netherlands: https://www.scp.nl/Onderzoek/Lopend_onderzoek/A_Z_alle_lopende_onderzoeken/Statusscores (date accessed 20 March 2017), 2016.
25. Liu JP, Hsueh HM, Hsieh E, Chen JJ. Tests for equivalence or non-inferiority for paired binary data. Stat Med 2002;21:231-45.
26. Fagerland MW, Lydersen S, Laake P. Recommended tests and confidence intervals for paired binomial proportions. Stat Med 2014;33:2850-75.
27. Alonzo TA, Braun TM, Moskowitz CS. Small sample estimation of relative accuracy for binary screening tests. Stat Med 2004;23:21-34.
28. De Girolamo G, Goldoni CA, Corradini R, Giuliani O, Falcini F, Sassoli De'Bianchi P, et al. Ambient temperature and FIT performance in the Emilia-Romagna colorectal cancer screening programme. J Med Screen 2016;23:186-191.
29. Bevolkingsonderzoek darmkanker: Monitor 2016 (Dutch). http://www.rivm.nl/Documenten_en_publicaties/Algemeen_Actueel/Uitgaven/Preventie_Ziekte_Zorg/Darmkankerscreening/Landelijke_Monitoring_en_Evaluatie_Bevolkingsonderzoek_Darmkanker.
30. Kapidzic A, van Roon AH, van Leerdam ME, van Vuuren AJ, van Ballegooijen M, Lansdorp-Vogelaar I, et al. Attendance and diagnostic yield of repeated two-sample faecal immunochemical test screening for colorectal cancer. Gut 2017;66:118-123.
31. Santare D, Kojalo I, Huttunen T, Rikacovs S, Rucevskis P, Boka V, et al. Improving uptake of screening for colorectal cancer: a study on invitation strategies and different test kit use. Eur J Gastroenterol Hepatol 2015;27:536-43.
32. Council of the European Union. Council Recommendation of 2 December 2003 on cancer screening (2003/878/EC). Off J Eur Union 2003:34-38.
33. Allison JE. The best screening test for colorectal cancer is the one that gets done well. Gastrointest Endosc 2010;71:342-5.
34. Bell KJ, Bossuyt P, Glasziou P, Irwig L. Assessment of changes to screening programmes: why randomisation is important. Bmj 2015;350:h1566.

APPENDIX

Appendix 1a. FOB-Gold Standard for Fecal Immunochemical Tests for hemoglobin (FITTER) checklist

Topic	Item	Documentation
Specimen collection and handling	Name of specimen collection device and supplier (address).	Name collection device: FOB-Gold Supplier: FOB-Gold, Sentinel Diagnostics, Via Robert Koch, 2 - 20152 Milan – Italy.
	Description of specimen collection device (vial with probe/stick, card, other).	Round tube with collection stick immerged in a preservative solution
	Description of specimens used if an *in vivo* study (single or pooled faeces, artificial matrix with added blood, etc).	Single human faeces sample.
	Details of fecal collection method (sampling technique and number of samples).	Ribbed section of the sampling stick is dipped in four different parts of the stool.
	Who collected the specimens from the samples (patient, technician, etc).	Participant
	Number of fecal specimens used in the study (single, pooled, individual patient faeces).	Single sample of individual patient faeces
	Mean mass of faeces collected.*	10 mg
	Volume of buffer into which specimen is taken by probe, applicator stick or card.*	1.7 mL
	Time and storage conditions of fecal specimen from "passing" to sampling, including time and temperature (median and range).	Analysis took place at same day of arrival (<24 hours) of the FIT in the lab and the FIT was kept by ambient air temperature.
	Time and storage of collection devices from specimen collection to analysis, including time and temperature (median and range). A concise description of process from collection to analysis is recommended.	Participants were asked to post the faeces samples within 24 hours after collection and keep the sample in the refrigerator. The date of sample collection is noted. FIT was transported and analyzed by ambient air temperature. All samples were analyzed within 5 days after collection.
Analysis		
	Name of analyzer, model, supplier (address), number of systems if more than one used.	Bio Majesty JCA-BM6010/C, serial number CA 1401000690069. Supplier: Sysmex Nederland B.V. Ecustraat 11 4879NP Etten Leur

Appendix 1a. (continued)

Topic	Item	Documentation
	Number of times each sample was analyzed.	Single or twice. If first analysis resulted in 'no results' analysis was repeated.
	Analytical working range* and whether samples outside this range were diluted (factor) and re-assayed.	0.4-797.2 ng/mL. Client samples outside this linearity range were not diluted.
	Source of calibrator(s) (supplier with address), number of calibrator(s), how concentrations were assigned* and details of calibration process including frequency.	Calibrator supplier: Sentinel Diagnostics, Via Robert Koch, 2 - 20152 Milan - Italy. Calibrator levels: 6 Standard calibration is performed with every reagent and calibrator lot number change.
	Analytical imprecision*, ideally with number of samples analyzed, concentrations, and mean, SD and CV.	Prior to the go-live a CLSI EP5A2 protocol was performed on all three Sentinel controls (low-mid-high) to verify the imprecision specifications as stated in the tender requirements of the colon cancer screening program. CLSI EP5A2 results:
		Sentinel LOW 50 ng/mL Lot number control 30004/A0546 SD with-in (calculated) = 4.04 SD with-in (claim) = 5.00 (10% of 50 ng/mL) (User variance/claim variance)*freedom degrees = 26.13 Critical Chi-square value = 55.76 Claim accepted? YES
		SD total (calculated) = 5.12 SD total (claim) = 7.50 (15% of 50 ng/mL) (User variance/claim variance)*freedom degrees = 25.63 Critical Chi-square value = 73.03 Claim accepted? YES
		Sentinel MID 71 ng/mL Lot number control 30004/A0551 SD with-in (calculated) = 3.69 SD with-in (claim) = 7.10 (10% of 71 ng/mL) (User variance/claim variance)*freedom degrees = 10.82 Critical Chi-square value = 55.76 Claim accepted? YES

Appendix 1a. (continued)

Topic	Item	Documentation
		SD total (calculated) = 5.49 SD total (claim) = 10.60 (15% of 71 ng/mL) (User variance/claim variance)*freedom degrees = 11.98 Critical Chi-square value = 61.66 Claim accepted? YES
		Sentinel HIGH 312 ng/mL Lot number control 30004/A0552 SD with-in (calculated) = 4.81 SD with-in (claim) = 31.20 (10% of 312 ng/mL) (User variance/claim variance)*freedom degrees = 0.95 Critical Chi-square value = 55.76 Claim accepted? YES
		SD total (calculated) = 7.41 SD total (claim) = 46.80 (15% of 312 ng/mL) (User variance/claim variance)*freedom degrees = 1.08 Critical Chi-square value = 59.30 Claim accepted? YES
Quality management	Source (address) or description of internal quality control materials, number of controls, assigned target concentrations and ranges, how target concentrations were assigned, rules used for acceptance and rejection of analytical runs.	3 rounds of control before running daily analyses were done and 3 rounds after, conform Sentinel's quality rules. If 2 out of 3 controls are within the range, analytical runs are accepted. Apart from the Sentinel controls, a mid-daily run of control conform SKML (SKML CFB, Mercator 1, Toernooiveld 214, NL-6525 EC, Nijmegen, The Netherlands) is performed, every other day a high run or low run: SKML low: 212 ng/mL ->5.05% SKML high: 510 ng/mL ->3.10% If the control is not right, controls are being repeated, if not right after multiple control rounds, de clinical chemist is consulted.
	Participation in external quality assessment schemes: (name and address of scheme), frequency of challenges, performance attained.	Participation in external quality assessments of SKML (foundation of quality control of medical laboratory diagnostics) following a fixed schedule. Assessment results are monitored by the national functionary iFOBT.

Appendix 1a. (continued)

Topic	Item	Documentation
	Accreditation held by the analytical facility (address).	Accreditation by CCKL, Mariaplaats 21-D, 3511 LK UTRECHT
	The number, training and expertise of the persons performing the analyses and recording the results	7 trained technician's
Result handling	Mode of collection of data- manual recording or via automatic download to IT system, single or double reading	Results are automatically uploaded to the Colonissystem, after authorization by the laboratory analyst the results are uploaded to the screening IT system ScreenIT.
	Units used, with conversions to µg Hb/g faeces if ng Hb/mL used.	In analyzing and reporting results ng Hb/mL was used. For reporting in publications this is converted to µg Hb/g faeces.[2]
	Cut-off concentration(s) if used and explanation of how assigned locally or by manufacturer	Positive: ≥88 ng Hb/mL. This was locally assigned by researchers and approved by the Ministry of Health.
	Were the analysts blinded (masked) to the results of the reference investigation and other clinical information?	Yes

Apendix 1b. OC-Sensor Standard for Fecal Immunochemical Tests for hemoglobin (FITTER) checklist

Topic	Item	Documentation
Specimen collection and handling	Name of specimen collection device and supplier (address).	Name collection device: S-bottle OC-Auto sampling bottle 3 Eiken Chemical Co LTD, 4-19-9 Taito, Taito-ku, Tokyo, 110-8408, Japan
	Description of specimen collection device (vial with probe/stick, card, other).	Flat tube with collection stick immerged in a preservative solution
	Description of specimens used if an *in vivo* study (single or pooled faeces, artificial matrix with added blood, etc).	Single human faeces sample
	Details of fecal collection method (sampling technique and number of samples).	Ribbed section of the sampling stick is striked 4 times through the stool.
	Who collected the specimens from the samples (patient, technician, etc).	Participant

Apendix 1b. (continued)

Topic	Item	Documentation
	Number of fecal specimens used in the study (single, pooled, individual patient faeces).	Single sample of individual patient faeces
	Mean mass of faeces collected.*	10 mg
	Volume of buffer into which specimen is taken by probe, applicator stick or card.*	2 mL
	Time and storage conditions of fecal specimen from "passing" to sampling, including time and temperature (median and range).	Analysis took place at same day of arrival of the FIT in the lab and the FIT was kept by ambient air temperature.
	Time and storage of collection devices from specimen collection to analysis, including time and temperature (median and range). A concise description of process from collection to analysis is recommended.	Participants were asked to post the faeces samples within 24 hours after collection and keep the sample in the refrigerator. The date of sample collection is noted. FIT was transported and analyzed by ambient air temperature. All samples were analyzed within 5 days after collection.
Analysis		
	Name of analyzer, model, supplier (address), number of systems if more than one used.	Diana OC Sensor, serial number SN N00738. Supplier: Eiken Chemical Co LTD, 4-19-9 Taito, Taito-ku, Tokyo, 110-8408, Japan
	Number of times each sample was analyzed.	Single or twice. If first analysis resulted in 'no results' analysis was repeated.
	Analytical working range* and whether samples outside this range were diluted (factor) and re-assayed.	50- 1000 ng/mL. Samples were not diluted.
	Source of calibrator(s) (supplier with address), number of calibrator(s), how concentrations were assigned* and details of calibration process including frequency.	Supplier: Eiken Chemical Co LTD, 4-19-9 Taito, Taito-ku, Tokyo, 110-8408, Japan. 1 calibrator measuring 6 dilutions. Calibration was done before start of the study. The same lot number was used during the study period. Control low: range 120 and 173 ng Hb/mL, lot number 5Z017 Control high: range 379 and 513 ng Hb/mL, lot number 5Z017 If all values were within the margins, the test samples were run for analysis.

Apendix 1b. (continued)

Topic	Item	Documentation
	Analytical imprecision*, ideally with number of samples analyzed, concentrations, and mean, SD and CV.	In total, 20 measurements were done for each control in 5 days. Each day 2 control were run once in the morning and evening. Control low: range concentrations measured: 148-178 ng Hb/ mL, mean 159.2 ng Hb/ mL, S total measured: 9.2 ng Hb/ mL (claim S total 15), total precision CV 5.8%. S within 9.5 (claim S total 10), within run precision CV: 6.0%. Control high: range concentrations measured: 389-512 ng Hb/ mL, mean 468.3 ng Hb/ mL, S total measured: 4.9 ng Hb/ mL (claim S total 32.8), total precision CV 1.0%. S within 20.0 (claim S total 23.4), within run precision CV: 4.3%.
Quality management	Source (address) or description of internal quality control materials, number of controls, assigned target concentrations and ranges, how target concentrations were assigned, rules used for acceptance and rejection of analytical runs.	Daily two rounds of control were done before analytical runs. Both controls should be right before start of analysis. If still not right calibration should follow (this has not happened during the study).
	Participation in external quality assessment schemes: (name and address of scheme), frequency of challenges, performance attained.	Not applicable for study period.
	Accreditation held by the analytical facility (address).	Accreditation by CCKL, Mariaplaats 21-D, 3511 LK UTRECHT
	The number, training and expertise of the persons performing the analyses and recording the results	4 trained technician's
Result handling	Mode of collection of data- manual recording or via automatic download to IT system, single or double reading	Data from analyzer were uploaded by usb-stick to the automatic IT system.
	Units used, with conversions to µg Hb/g faeces if ng Hb/ mL used.	In analyzing and reporting results ng Hb/ mL was used. For reporting in publications this is converted to µg Hb/g faeces.[2]
	Cut-off concentration(s) if used and explanation of how assigned locally or by manufacturer	Positive: ≥75 ng Hb/ mL. This was assigned by researchers and approved by the Ministry of Health.

Apendix 1b. (continued)

Topic	Item	Documentation
	Were the analysts blinded (masked) to the results of the reference investigation and other clinical information?	Yes

[1.] Fraser CG, Allison JE, Young GP, Halloran SP, Seaman HE. Improving the reporting of evaluations of fecal immunochemical tests for haemoglobin: the FITTER standard and checklist. Eur J Cancer Prev 2015; 24(1): 24-6.
[2.] 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): 810-14.