Title: Overestimated Sensitivity of Fecal Immunochemical Tests in Screening Cohorts With Registry-Based Follow-up.

Authors: Brenner H, Gies A, Selby K

Journal: The American journal of gastroenterology

Year: 2019 Nov

Issue: 114

Volume: 11

Pages: 1795-1801

DOI: 10.14309/ajg.0000000000000412

Overestimated sensitivity of fecal immunochemical tests

in screening cohorts with registry-based follow-up

Short title: Diagnostic performance of FITs

Hermann Brenner, MD[1,2,3] Anton Gies, pharmacst[2,4] Kevin Selby, MD, MAS[5]

[1] Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany

[2] Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany

[3] German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany

[4] Medical Faculty Heidelberg, University of Heidelberg, Heidelberg, Germany

[5] Center for Primary Care and Public Health (unisanté), University of Lausanne, Lausanne, Switzerland

Abbreviations: AA, advanced adenoma; AN, advanced neoplasm; CRC, colorectal cancer; FIT, fecal immunochemical test

Address for correspondence:
Hermann Brenner, Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, Im Neuenheimer Feld 581, 69120 Heidelberg, Germany
Phone ++49-6221-421300, Fax ++49-6221-421302

Guarantor of the article: Hermann Brenner

Specific author contributions: HB designed the study, carried out the statistical analyses and wrote the manuscript. AG and KS contributed important intellectual content to critical revision of the manuscript. All authors approved the final submitted version.

Financial support: This work was supported in part by a grant from the German Federal Ministry of Education and Research (grant number 01GL1712). The sponsor did not have any role in the study design, in the collection, analysis, and interpretation of data.

Potential competeing interests: The authors declare no conflict of interest.

Word Count: 2870

Study Highlights

WHAT IS KNOWN

- Reported estimates of sensitivity of faecal immunochemical tests (FITs) for detecting colorectal cancer varied widely.
- Very high, seemingly very precise estimates were reported from registry-based follow-up-studies of FIT participants.

WHAT IS NEW HERE

- It is shown that such registry-based follow-up-studies provide very accurate estimates of specificity.
- Sensitivities derived from such studies are shown to be subject to potential severe overestimation, however.
- These results may prevent unrealistic expectations with respect to sensitivity of FITs for CRC detection.

Abstract

Objectives. Several recent studies have reported very high estimates of sensitivity and specificity of fecal immunochemical tests (FITs) at seemingly high levels of precision using registry-based follow-up of participants in very large FIT-based screening programs. We aimed to assess the validity of estimates of diagnostic performance parameters derived by this indirect approach.

Methods. We modeled expected values of sensitivity and specificity of CRC detection in studies using the indirect approach and their deviation from true values under a broad range of plausible assumptions, and we compared these expected values with recently reported estimates of FIT sensitivity and specificity from such studies.

Results. Using a sensitivity of 75% and specificity of 93.6% (from studies using a direct approach, i.e., colonoscopy follow-up of all participants), the indirect approach would be expected to yield sensitivities between 84.5% and 91.1% and specificities between 93.4% and 93.6% under a range of realistic assumptions regarding colonoscopic follow-up rates of positive FITs and clinical manifestation rates of preclinical CRC.

Conclusions. Very high sensitivities of FITs recently reported with seemingly very high levels of precision by several large scale registry-based studies, which are in line with expected results based on our model calculations, are likely to be strongly overestimated and need to be interpreted with due caution.

Key words: colorectal cancer, diagnostic performance, fecal occult blood test, screening

Introduction

Fecal immunochemical tests (FITs) for hemoglobin are used for colorectal cancer (CRC) screening in an increasing number of countries.[1] There have been two major approaches to estimate diagnostic performance of FITs in screening settings:[2]

(i) „Direct approach“: Ascertainment of presence or absence of CRC at the time of FIT screening by direct comparison with results of screening colonoscopy in all study particpants regardless of FIT results.

(ii) „Indirect approach“: Ascertainment of presumed presence or absence of CRC at the time of FIT screening by colonoscopy follow-up of FIT positive results only, and by registry-based follow-up, such as record linkage with cancer registries, medical records or insurance claims, among the vast majority of participants with a negative FIT result or a positive FIT not followed by colonoscopy.

The direct approach is conceptually straightforward, but requires large cohorts of participants undergoing screening colonoscopy in whom FIT is additionally conducted prior to bowel preparation. The indirect approach is increasingly used in settings in which FIT screening is established and reliable CRC identification is possible by record linkage with routinely collected data. A systematic review and meta-analysis of the pre-2014 literature identified 19 studies on diagnostic accuracy of FITs.[2] Twelve studies using the direct approach included 239 CRC cases, seven studies using the indirect approach[3-9] included 198 CRC cases. Although summary estimates of specificity were similar for both types of studies, the summary estimate of sensitvity was substantially higher (87%) for studies using the indirect approach than for those using the direct approach (71%). The overall summary estimate of both types of studies combined, reported as the main result of the meta-analysis, was 79%.[2] In recent years, very large studies using the indirect approach have been published from established FIT-based screening programs, which in sum included almost 5000 CRC cases, so that evidence for FIT sensitivity now predominantly comes from those studies.[10-13]

The indirect approach relies on two key assumptions, i.e. that

(i) all CRC cases detected within 2 years derive from preclinical CRCs that are already present at the time of FIT screening, and

(ii) all preclinical CRCs present at the time of FIT screening become clinically manifest or otherwise detected within the follow-up period (typically 2 years).

Violations of the first assumption are likely to be small, given the rather low transition rates from advanced adenomas, the most common precursors of CRC, to preclinical CRC, which have been estimated in the order of 2.5 to 5.6% per year[14] and the rather long mean sojourn time of preclinical CRC (i.e., the mean time CRC remains undetected in the absence of screening), which has been estimated in the order of 3 to 7 years[15-18]). This rather long sojourn time implies potential major violation of assumption (ii), however, and it is unclear to what extent such violation, also known as „verification bias“, may affect estimates of diagnostic performance of FITs. In this paper, we aimed to evaluate the validity of estimates of diagnostic performance of FITs for CRC detection derived by the indirect approach using model calculations and comparisons of their results with recently reported estimates form large scale registry-based studies.

Methods

Model set-up and model parameters

We carried out model calculations to estimate the expected apparent sensitivity and specificity of CRC detection compared to the true sensitivity and specificity in a cohort of participants of FIT screening who are subsequently followed with respect to CRC detection within two years. The model takes into account prevalence of CRC and its most common precursor, advanced adenoma (AA), at the time of FIT screening, the true sensitivity of FIT for CRC and AA, the true specificity for absence of any advanced neoplasm (AN, i.e., either CRC or AA), the colonoscopy uptake rate following a positive FIT result, and the 2-year transition rate from preclinical CRC to clinically manifest (diagnosed) CRC. Mortality is not considered given the relatively low 2-year mortality in the screening age population. **Table 1** provides an overview of the model parameters, the assumed true parameter values for the base case analyses and for sensitivity analyses, and references that were used as source for choosing those values. Note that sensitivity and specificity in our calculations refer to application of a single FIT, the most commonly reported parameters of diagnostic performance, rather than sensitivity and specificity for repeated FITs that are commonly offered in screening programs.

Derivation of expected apparent prevalence of CRC and diagnostic performance of FITs

In the following we derive the apparent prevalence of CRC and diagnostic performance of FITs expected with the indirect approach and illustrate the derivation with a numerical example of a cohort of 100,000 FIT screening participants (**Figure 1**).

Using the notation given in **Table 1**, the apparent prevalence of CRC, i.e., the denominator for the apparent sensitivity of FIT for detecting CRC expected from the indirect approach, based on 2-year follow-up of the cohort, includes the following three components:

A. FIT positive CRC cases verified through subsequent colonoscopy. This component is given by

$$A = P_{CRC} \times SE_{CRC} \times CS_{FU}.$$

B. FIT positive CRC cases that are not followed up by colonoscopy but become clinically manifest during the 2-year follow-up. This component is given by

$$B = P_{CRC} \times SE_{CRC} \times (1 - CS_{FU}) \times CM_{CRC\_2y}.$$

C. FIT negative CRC cases that become clinically manifest during the 2-year follow-up. This component is given by

$$C = P_{CRC} \times (1 - SE_{CRC}) \times CM_{CRC\_2y}.$$

The expected apparent sensitivity is given by the FIT positive components of the apparent CRC prevalence, i.e.

$$(A + B) / (A + B + C).$$

Note that our model calculations follow the assumption implicitly made in the indirect approach that all CRC cases detected within 2 years derive from preclinical CRCs are already present at the time of FIT screening. However, whereas the indirect approach implicitly additionally assumes that all preclinical CRCs present at the time of FIT screening become clinically manifest and detected within 2 years, our model allows for less than complete CRC manifestation and detection within this restricted time window. A simplifying assumption of our model is that the clinical manifestation rate within 2 years (whose relevance will be critically discussed below) is the same for FIT positive and FIT negative preclinical CRC cases that are not followed up by colonoscopy.

The denominator of the apparent specificity, i.e. of the probability of a negative FIT result in the (apparent) absence of CRC, is given as 1 minus the apparent CRC prevalence, i.e.

$$D = 1 – (A + B + C).$$

The numerator of the apparent specificity is expected to be

$$E = P_{CRC} \times (1 - SE_{CRC}) \times (1 - CM_{CRC\_2y}) + P_{AA} \times (1 - SE_{AA}) + (1 - P_{CRC} - P_{AA}) \times SP_{noAN}.$$

Hence, the apparent specificity for „no CRC“ is expected to be E / D. The true specificity is given by

$$(P_{AA} \times (1 - SE_{AA}) + (1 - P_{CRC} - P_{AA}) \times SP_{noAN}) / (1 - P_{CRC}).$$

Base case analyses and sensitivity analyses

We first derived the true and expected apparent prevalence, sensitivity, specificity, positive and negative predictive value using the base case parameter values shown in **Table 1**. We then repeated the calculations assuming the alternative values for each of the parameters. In order to illustrate the specific impact of each parameter, variations were done separately for each parameter (i.e., for one parameter at a time).

Results

Figure 1 shows the expected numbers of participants diagnosed with CRC within 2 years in a hypothetical cohort of 100,000 FIT participants, based on our model using the base case parameter values shown in **Table 1**.

Derivation of the true and apparent sensitivity and specificity of CRC detection for this hypothetical cohort is illustrated in **Table 2**. Overall, 420+35+58 = 513 out of 700 participants with CRC would be diagnosed with the disease within two years. Of the 513 detected CRC cases, 420+35=455 would have a positive FIT result, yielding an apparent sensitivity of 455/513 = 88.7%. This estimate exceeds the assumed true sensitivity, which equals 525/700 = 75%, by 13.7 percentage points. The apparent specificity would be 93.5%, i.e., very close to the true specificity of 93.6%.

As shown in **Table 3**, the sensitivity of FIT for detecting CRC is consistently overestimated in all of the assessed scenarios. For the scenarios assuming a true sensitivity of 75%, observed sensitivities are expected to range from 84.7% to 91.1%. By contrast, the apparent specificity is very close to the true specificity in all of the assessed scenarios, with underestimation of less than or equal to 0.2 percentage points in all scenarios.

Discussion

In this paper, we provide a thorough analysis of the use and limitations of a commonly and increasingly employed indirect approach of estimating diagnostic performance of FITs by record linkage of large cohorts of FIT participants with routinely collected follow-up data. Although this approach yields accurate estimates of specificity under a broad range of assumptions, estimates of sensitivity of detecting CRC are prone to strong overestimation. Very high sensitivities reported with apparently very high precision in recent registry based follow-up studies of large cohorts of FIT participants therefore need to be interpreted with utmost caution.

The indirect approach would be expected to provide valid estimates of sensitivity if all preclinical CRCs present at the time of FIT application became clinically manifest or were otherwise detected during the commonly employed one- to two-year window of follow-up. However, this assumption, whose violation is the key driver of overestimtion of sensitivity, is problematic. Estimates of mean sojourn time of CRC have been consistently in the range from 3 to 7 years[15-18] which translates to the range of cumulative 2-year clinical manifestation rates from 1 – exp(-2/7) = 25% to 1 – exp(-2/3) = 49% that was covered by our sensitivity analyses. Although the majority of studies using the indirect approach were based on two years of follow-up, some recent large-scale studies included a 1-year follow-up only.[10,11] Clinical manifestation rates are expected to be substantially lower during such a short time window and overestimation of sensitivity is therefore expected to be of particular concern in such studies.

Given the still rather limited number of studies providing direct estimates of diagnostic performance of FIT in screening settings and given the rather limited number of CRC cases in those studies (ranging up to 79 CRC cases only),[21] studies using the indirect approach have accounted for a substantial and rapidly increasing proportion of available “evidence” of

diagnostic performance for CRC detection of FITs. For example, in the systematic review and meta-analysis of the pre-2014 literature by Lee et al,[2] studies using an indirect approach contributed larger numbers of participants and almost the same numbers of CRC cases to the derivation of summary estimates of FIT sensitivity and specificity compared to studies using a direct approach (198 and 239 CRC cases, respectively). In concordance with the results of the current analysis, summary estimates of specificity had been very similar for both types of studies whereas summary estimates of sensitivity were much higher for studies using the indirect approach (87%) compared to studies using a direct approach (71%).

In recent years, a rapidly increasing number of registry-based studies have used the indirect approach to estimate FIT sensitivity and specificity, some of which included very large numbers of FIT participants and CRC cases that by far exceeded corresponding numbers from studies using a direct colonoscopy-controlled approach. In particular, four recent studies from the US and Taiwan each included more than 300,000 participants of FIT based screening programs (**Table 4**). With the exception of one study from Taiwan,[11] participants were between 50 and 69-75 years old, and all four studies included slightly more women than men. Between 645 and 2005 CRC cases were identified through screening program records and cancer registry-based follow-up over one or two years from participation in FIT based screening.[10-13] Together these studies included almost 5,000 CRC cases. All four studies reported sensitivity and specificity for a FIT cutoff at 20 µg hemoglobin per g feces. Although estimates of specificity were rather similar in these studies (and comparable to estimates from studies using the direct approach), ranging from 92.6%[13] to 96.2%,[12] there were large differences in reported sensitivity, ranging from 74.3% (95% CI 71.8-76.7%) to 93.3% (95% CI 91.6-94.9%),[10-13] despite the use of quantitative FITs (OC Sensor, OC FIT-CHEK) with the same positivity threshold (20 µg Hb/g feces). Given the size of the studies and the narrow confidence intervals around the point estimates of sensitivity, these differences are far beyond what can be explained by chance. Reported sensitivities were particularly high in the two studies based on 1-year follow-up only (84.5% and 93.3%).[10,11]

The results of our model calculations suggest that these apparently very high sensitivities might mainly reflect even less complete clinical manifestation of preclinical CRC during one year of follow-up only, thereby accentuating the expected overestimation of sensitivity.

A major advantage of the indirect approach to estimate diagnostic performance of FITs and the main reason for its increasing popularity is that it can be applied without the major extra efforts and logistics of study related data collection required for studies using the direct approach. Participants in studies using the indirect approach may also better represent real life FIT screening populations. In settings where routine data from FIT-based screening programs are readily available and can be linked to routine data sources providing reliable information on new diagnoses of cancer, such as cancer registries, the indirect approach can therefore be a very efficient and highly valuable tool for valid estimation of FIT specificity in a real life stetting even though estimates of sensitivity must be regarded with utmost caution. Follow-up intervals of at least 2 years seem to be required in order to limit overestimation of sensitivity. Although missed CRCs (the source of this overestimation) would be of less concern with even longer follow-up periods, longer follow-up periods might also be problematic due to the risk of including fast growing CRCs that might not have been present at the time of the initial FIT.

Apart from providing estimates of sensitivity for detecting CRCs, provision of estimates of sensitivity for detection of precancerous lesions is a major contribution of studies evaluating diagnostic performance of FITs using the direct approach.[21-24] Since most precancerous lesions remain undetected, a further limitation of studies using the indirect approach is that they cannot provide meaningful sensitivity estimates for detection of those lesions.

Although our analyses focused on estimates sensitivity and specificity, additional parameters of diagnostic performance from the indirect approach can easily be derived from our results. For example, relative overestimation of positive likelihhod ratios, which are derived as

sensitivity divided by (1-specificity), is expected to be very close to relative overestimation of sensitivity that can be derived from results shown in Table 3, given that the expected bias in estimates of specificity is close to negligible. For the same reason, underestimation of negative likelihood ratios, which are derived as (1-sensitivity) divided by specificity, is expected to be very close to underestimation of (1-sensitivity). Despite potential major overestimation of sensitivity, the indirect approach might still provide valid estimates of the positive predictive value (PPV), a most relevant diagnostic performance parameter for clinical practice, provided that colonoscopy follow-up rates are the same for FIT positive participants with CRC, advanced adenomas or no advanced neoplasms. For example, using the numerical illustration provided in Figure 1, the apparent PPV in a study using the indirect approach could be derived from colonoscopy follow-up data of positive FITs as 420/(420+1400+3692) = 7.6%, exactly the same as the true positive predictive value of 525/(525+1750+4615) = 7.6%, assuming that the “equal colonoscopy follow-up assumption” holds. In practice, however, colonoscopy follow-up rates might by higher among FIT positive carriers of CRC than among those without CRC, as quantitative FIT results are expected to be substantially higher in the former, which may also lead to some overestimation PPVs.

In summary, our model calculations provide important clues as to the use and limitations of a commonly and increasingly employed indirect approach of estimating diagnostic performance of FITs by registry-based follow-up. Although the indirect approach is very useful to provide reliable estimates of FIT specificity, sensitivity for detecting precancerous lesions cannot be estimated and sensitivity for detecting CRCs may be seriously overestimated. Such overestimation is of particular concern in studies with relatively short follow-up of FIT participants. In order to avoid unrealistic expectations of FIT performance in planning of screening programs, decisions on the choice of screening tests and communication of diagnostic properties of FIT to potential screening participants, estimates of sensitivity should be based on studies using the direct approach, i.e. studies with colonoscopic follow-up of all participants. The previously reported summary estimate of

sensitivity from such studies, 71% (95% CI 58-81%),[2] likely reflects true sensitivity of FITs much better than the apparently much higher and much more “precise” estimates reported by the recent large scale studies using the indirect approach. Given that numbers of CRC cases are typically much larger in the latter studies, major caution is also warranted when summary estimates of sensitivity are derived from studies using the direct and indirect approach, as these summary estimates will be heavily and increasingly dominated by the apparently more precise but much more bias-prone estimates from the studies using the indirect approach.

References

1. Schreuders EH, Ruco A, Rabeneck L,et al. Colorectal cancer screening: a global overview of existing programmes. Gut 2015; 64:1637-49.
2. Lee JK, Liles EG, Bent S, et al. Accuracy of fecal immunochemical tests for colorectal cancer: systematic review and meta-analysis. Ann Intern Med 2014; 160: 171.
3. Nakama H, Kamijo N, Abdul Fattah AS, et al. Validity of immunological faecal occult blood screening for colorectal cancer: a follow up study. J Med Screen 1996; 3: 63–5.
4. Allison JE, Tekawa IS, Ransom LJ, et al. A comparison of fecal occult-blood tests for colorectal-cancer screening. N Engl J Med 1996; 334:155–9.
5. Itoh M, Takahashi K, Nishida H, et al. Estimation of the optimal cut off point in a new immunological faecal occult blood test in a corporate colorectal cancer screening programme. J Med Screen 1996; 3: 66–71.
6. Launoy GD, Bertrand HJ, Berchi C, et al. 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: 493–6.
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: 1462–70.
8. Parra-Blanco A, Gimeno-Garcia AZ, Quintero E, et al. Diagnostic accuracy of immunochemical versus guaiac faecal occult blood tests for colorectal cancer screening. J Gastroenterol 2010; 45: 703–712.
9. 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.
10. 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: 456-63.
11. Chen CH, Wen CP, Tsai MK. Fecal immunochemical test for colorectal cancer from a prospective cohort with 513,283 individuals: Providing detailed number needed to scope (NNS) before colonoscopy. Medicine 2016; 95: e4414.
12. Chen SL-S, Hsu C-Y, Yen AM-F, et al. Demand for Colonoscopy in Colorectal Cancer Screening Using a Quantitative Fecal Immunochemical Test and Age-Sex-Specific Thresholds for Test Positivity. Cancer Epidemiol Biomarkers Prev 2018; 27: 704-10.
13. Selby K, Jensen CD, Lee JK, et al. Influence of Varying Quantitative Fecal Immunochemical Test Positivity Thresholds on Colorectal Cancer Detection: A Community-Based Cohort Study. Ann Intern Med 2018; 169: 439-47.
14. Brenner H, Altenhofen L, Stock C, et al. Natural history of colorectal adenomas: birth cohort analysis among 3.6 million participants of screening colonoscopy. Cancer Epidemiol Biomarkers Prev 2013; 22: 1043-51.
15. Launoy G, Smith TC, Duffy SW, et al. Colorectal cancer mass-screening: estimation of fecal occult blood test sensitivity, taking into account cancer mean sojourn time. Int J Cancer 1997; 73: 220–4.
16. Prevost TC, Launoy G, Duffy SW, et al. Estimating sensitivity and sojourn time in screening for colorectal cancer: a comparison of statistical approaches. Am J Epidemiol 1998; 148: 609–19.

17. Lansdorp-Vogelaar I, van Ballegooijen M, Boer R, et al. A novel hypothesis on the sensitivity of the fecal occult blood test: results of a joint analysis of 3 randomized controlled trials. Cancer 2009; 115: 2410–9.

18. Brenner H, Altenhofen L, Katalinic A, et al. Sojourn time of preclinical colorectal cancer by sex and age: estimates from the German national screening colonoscopy database. Am J Epidemiol 2011; 174: 1140-6.

19. Brenner H, Altenhofen L, Kretschmann J, et al. Trends in Adenoma Detection Rates During the First 10 Years of the German Screening Colonoscopy Program. Gastroenterology 2015; 149: 356-66.

20. Imperiale TF, Ransohoff DF, Itzkowitz SH, et al. Multitarget stool DNA testing for colorectal-cancer screening. N Engl J Med 2014; 371: 187-8.

21. Gies A, Bhardwaj M, Stock C, et al. Quantitative fecal immunochemical tests for colorectal cancer screening. Int J Cancer 2018; 143: 234-44.

22. 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(1):93-104.

23. Brenner H, Calderazzo S, Seufferlein T, Ludwig L, Dikopoulos N, Mangold J, Böck W, Stolz T, Eisenbach T, Block T, Kopp-Schneider A, Czock D, Tikk K. Effect of a Single Aspirin Dose Prior to Fecal Immunochemical Testing on Test Sensitivity for Detecting Advanced Colorectal Neoplasms: A Randomized Clinical Trial. JAMA 2019;7;321(17):1686-1692.

24. Imperiale TF, Gruber RN, Stump TE, Emmett TW, Monahan PO. Performance Characteristics of Fecal Immunochemical Tests for Colorectal Cancer and Advanced Adenomatous Polyps: A Systematic Review and Meta-analysis. Ann Intern Med 2019 Feb 26. [Epub ahead of print].

Table 1. Parameters used for the model calculations

Parameter		Values [%]		References
Notation	Explanation	Base case	Alternative	
P_{CRC}	Prevalence of CRC	0.7	0.4, 1.0	19, 20
P_{AA}	Prevalence of advanced adenoma	7	4, 10	19, 20
SE_{CRC}	True sensitivity of CRC detection	75	60, 90	2, 21
SE_{AA}	True sensitivity of AA detection	25	10, 40	21
SP_{noAN}	True specificity (no AN)	95	92, 98	21
CS_{FU}	Colonoscopic follow-up rate of positive FIT results	80	40, 60, 100	4-6, 10
CM_{CRC_2y}	2-year clinical manifestation rate of preclinical CRC*	33	25, 49	15-18

Abbreviation: CRC, colorectal cancer

* Estimates of mean sojourn time of CRC before clinical manifestation in references 16-19 have been consistently in the range from 3 to 7 years. We assumed a mean sojourn time of 5 years for the base case analysis and 3 and 7 years for the sensitivity analyses which translates to cumulative 2-year clinical manifestation rates of 1–exp(-2/5) = 33%, 1–exp(-2/3) = 49%, and 1–exp(-2/7) = 25%, respectively.

Table 2. True and apparent sensitivity and specificity of colorectal cancer detection for the hypothetical cohort shown in Figure 1.

Parameter	Estimate	Numerator	Denominator	Value
Sensitivity	True	420+35+70	700	75.0%
	Apparent	420+35	420+35+58	88.7%
Specificity	True	5250+87,685	100,000-700	93.6%
	Apparent	5250+87,685+117	100,000-(420+35+58)	93.5%

Table 3. True and apparent sensitivity and specificity of colorectal cancer detection in the base case scenario and sensitivity analyses

Scenario	Parameter [%]		Sensitivity [%]			Specificity [%]		
			True	Apparent	Difference	True	Apparent	Difference
Base case	See Table 1		75.0	88.7	+13.7	93.6	93.5	-0.1
Sensitivity analyses: varied parameter	P_{CRC}	0.4	75.0	88.7	+13.7	93.6	93.6	±0.0
		1.0	75.0	88.7	+13.7	93.6	93.5	-0.1
	P_{AA}	4	75.0	88.7	+13.7	94.2	94.1	-0.1
		10	75.0	88.7	+13.7	93.0	92.9	-0.1
	SE_{CRC}	60	60.0	79.8	+19.8	93.6	93.5	-0.1
		90	90.0	95.9	+ 5.9	93.6	93.5	-0.1
	SE_{AA}	10	75.0	88.7	+13.7	94.6	94.6	±0.0
		40	75.0	88.7	+13.7	92.5	92.5	±0.0
	SP_{noAN}	92	75.0	88.7	+13.7	90.8	90.7	-0.1
		98	75.0	88.7	+13.7	96.4	96.3	-0.1
	CS_{FU}	40	75.0	84.5	+ 9.5	93.6	93.4	-0.2
		60	75.0	86.9	+11.9	93.6	93.5	-0.1
		100	75.0	90.1	+15.1	93.6	93.6	±0.0
	CM_{CRC_2y}	25	75.0	91.1	+16.1	93.6	93.5	-0.1
		49	75.0	84.7	+ 9.7	93.6	93.5	-0.1

Abbreviations: CM_{CRC_2y}, 2-year clinical manifestation rate of preclinical colorectal cancer; COL_{FU}, colonoscopic follow-up rate of positive FIT results; CRC, colorectal cancer; SE_{AA}, true sensitivity of advanced adenoma detection; SE_{CRC}, true sensitivity of colorectal cancer detection; P_{AA}, prevalence of advanced adenoma; P_{CRC}, prevalence of colorectal cancer; SP_{noAN}, true specificity (no AN)

Table 4. Estimates of sensitivity and specificity of FITs reported by five recent large studies (>500 CRC cases) using the indirect approach

Authors, year	Country, study period	Study participants			FIT brand	Cutoff [µg/g]	Follow-up		CRC cases	Sensitivity (95% CI)	Specificity (95% CI)
		N	Age	Sex			Source	Years			
Chen et al, 2016[11]	Taiwan, 1994-2007	513,283	20+ (54%≥40)	48% men	OC Sensor	20	Cancer registry	1	921	93.3 (91.6-94.9)	96.0 (95.9-96.0)
Jensen et al, 2016[10]	US 2007-2013	323,349	50-70 (mean 58.5)	46% men	OC FIT-CHEK	20	Cancer registry	1	645	84.5 (81.5-87.1)[a]	95.2 (95.1-95.3)[a]
Chen et al, 2018[12]	Taiwan 2004-2009	723,113	50-69 (60% 50-59)	38% men	OC Sensor	20	Cancer registry	2	2005	78.7 (76.9-80.4)	96.2 (96.1-96.3)
Selby et al, 2018[13]	US 2013-2016	640,859	50-75 (87% 50-69)	47% men	OC FIT-CHEK	20[b]	Cancer registry	2	1245	74.3 (71.8-76.7)[c]	92.6 (92.5-92.6)[d]

Abbreviations: CRC, colorectal cancer; FIT, fecal immunochemical test

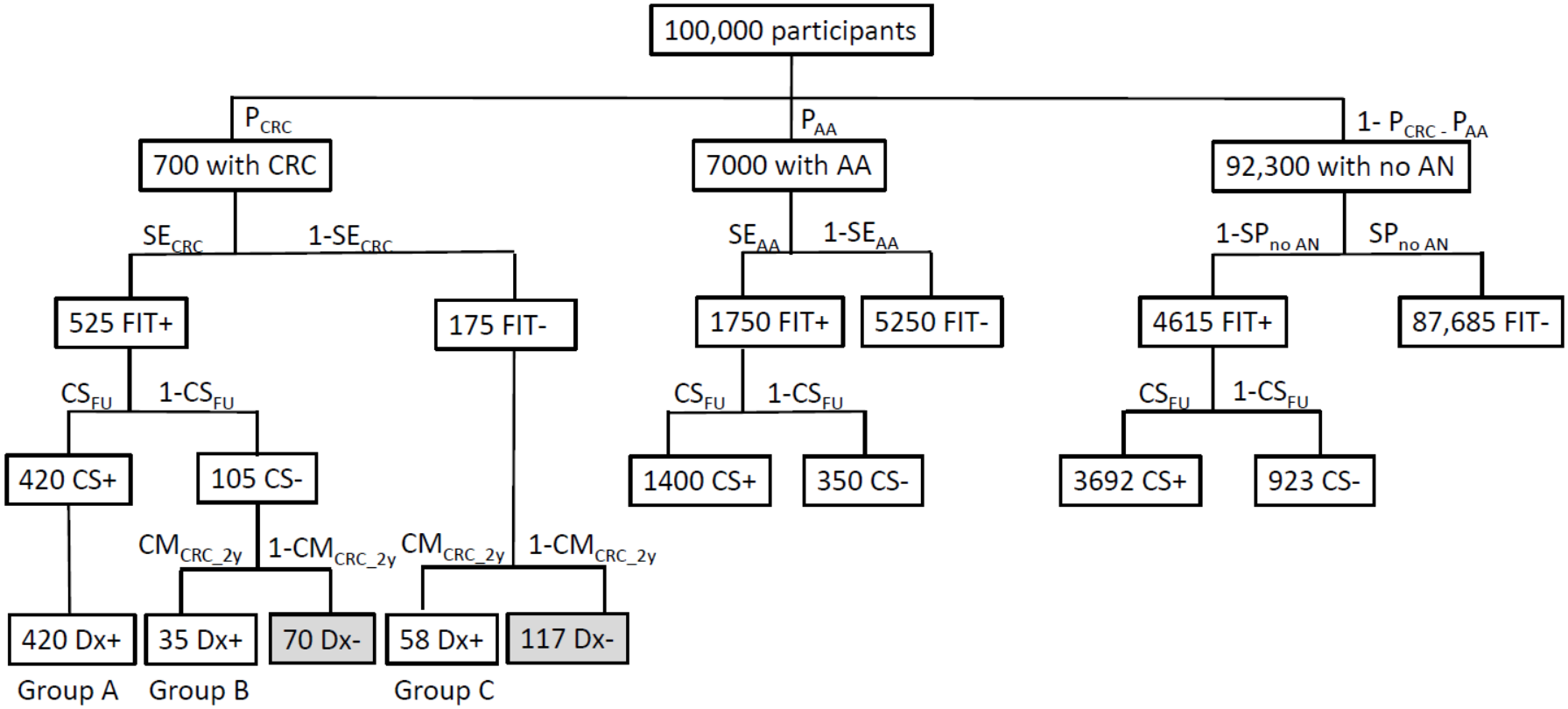
[a] 95% CI not reported by the authors but calculated from reported case numbers

[b] This study also evaluated other cutoffs, but results for cutoff 20 µg/g are listed only here for the sake of comparability with the other studies

[c] „Programmatic sensitivity“, defined as proportion of patients with a CRC diagnosis within 2 years of FIT screening who had a quantitative FIT result at baseline or during follow-up testing

[d] „Programmatic specificity“, defined as proportion of patients without a CRC diagnosis whose quantitative FIT results were all negative

Figure 1. Expected numbers of participants diagnosed with CRC within 2 years in a hypothetical cohort of 100,000 FIT screening participants, based on our model using the base case parameter values shown in Table 1. The grey shaded cells indicate the CRC cases missed by the indirect approach.



Abbreviations: AA, advanced adenoma; CM_{CRC_2y}, 2-year clinical manifestation rate of preclinical colorectal cancer; CRC, colorectal cancer; CS_{FU}, colonoscopic follow-up of positive FIT results; CS+, CS-: uptake of colonoscopy yes/no; Dx+, Dx-: diagnosis of colorectal cancer yes/no; FIT+/FIT-, positive/negative result of fecal immunochemical test; P_{AA} / P_{CRC}, prevalence of advanced adenoma / colorectal cancer; SE_{AA} / SE_{CRC}, sensitivity of detecting advanced adenoma / colorectal cancer; SP_{noAN}, specificity for absence of advanced neoplasm