

Open

Stage-Specific Sensitivity of Fecal Immunochemical Tests for Detecting Colorectal Cancer: Systematic Review and Meta-Analysis

Tobias Niedermaier, MPH, PhD^{1,2}, Yesilda Balavarca, PhD^{3,4} and Hermann Brenner, MD, MPH^{1,3,4}

- OBJECTIVE:** Fecal immunochemical tests (FITs) detect the majority of colorectal cancers (CRCs), but evidence for variation in sensitivity according to the CRC stage is sparse and has not yet been systematically synthesized. Thus, our objective was to systematically review and summarize evidence on the stage-specific sensitivity of FITs.
- METHODS:** We screened PubMed, Web of Science, Embase, and the Cochrane Library from inception to June 14, 2019, for English-language articles reporting on the stage-specific sensitivity of FIT for CRC detection using colonoscopy as a reference standard. Studies reporting stage-specific sensitivities and the specificity of FIT for CRC detection were included. Summary estimates of sensitivity according to the CRC stage and study setting (screening cohorts, symptomatic/diagnostic cohorts, and case-control studies) were derived from bivariate meta-analysis.
- RESULTS:** Forty-four studies (92,447 participants including 3,034 CRC cases) were included. Pooled stage-specific sensitivities were overall very similar but suffered from high levels of imprecision because of small case numbers when calculated separately for screening cohorts, symptomatic/diagnostic cohorts, and case-control studies. Pooled sensitivities (95% confidence intervals) for all studies combined were 73% (65%–79%) for stage-I-CRCs and 80% (74%–84%), 82% (77%–87%), and 79% (70%–86%) for the detection of CRC stages II, III, and IV, respectively. Even substantially larger variation was seen in sensitivity by T-stage, with summary estimates ranging from 40% (21%–64%) for T1 to 83% (68%–91%) for T3-CRC.
- DISCUSSION:** Although FITs detect 4 of 5 CRCs at stages II–IV, the substantially lower sensitivity for stage-I-CRC and, in particular, T1 CRC indicates both need and potential for further improvement in performance for the early detection of CRC.

SUPPLEMENTARY MATERIAL accompanies this paper at <http://links.lww.com/AJG/B321>, <http://links.lww.com/AJG/B322>, <http://links.lww.com/AJG/B319>, <http://links.lww.com/AJG/B320>, <http://links.lww.com/AJG/B318>

Am J Gastroenterol 2020;115:56–69. <https://doi.org/10.14309/ajg.000000000000465>

INTRODUCTION

Fecal immunochemical tests (FITs) are used for colorectal cancer (CRC) screening in a large and increasing number of countries (1). Multiple large-scale screening cohorts have demonstrated that FITs detect the majority of CRCs. In a meta-analysis from 2019, pooled sensitivity ranged from 71% (95% confidence interval [CI]: 56%–83%) to 91% (84%–95%), and specificity ranged from 90% to 95% (2). However, little is known on the sensitivity of FIT for detecting CRC at individual stages because many of the individual studies did not report stage-

specific sensitivities, and no meta-analysis of stage-specific sensitivities was conducted.

The ability of FITs to detect CRC at early stages is of particular relevance for their use in CRC screening, as chances of cure of CRC are dramatically higher when they are detected in earlier rather than later stages (3). Despite numerous studies reporting on stage-specific test characteristics of FIT (4–43), only one study to date (22) focused on this outcome. Given the small case numbers from the studies conducted among asymptomatic screening participants, obtaining reasonably precise stage-

¹Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany; ²Medical Faculty Heidelberg, University of Heidelberg, Heidelberg, Germany; ³Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany; ⁴German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany.

Correspondence: Tobias Niedermaier. E-mail: t.niedermaier@dkfz.de

Received March 18, 2019; accepted October 21, 2019; published online December 10, 2019

specific estimates of sensitivity from such studies is difficult if not impossible. Much larger case numbers have been included in the studies conducted among symptomatic participants who were diagnosed with CRC following colonoscopy for clarification of symptoms (symptomatic/diagnostic cohorts) or who were recruited after CRC diagnosis and compared with healthy controls in a clinical setting (case-control studies).

Although overall sensitivity is expected to be higher among these patient groups than among average-risk participants undergoing screening colonoscopy due to spectrum bias, stage-specific sensitivity may be comparable (44). In that case, the possibility of recruiting larger number of CRC patients in such settings might enable estimating stage-specific sensitivities of FIT irrespective of the study type at much higher levels of precision. Our aim was therefore to systematically review and compare the estimates of the stage-specific sensitivity of FIT for CRC detection from various types of studies and, if possible, summarize them to derive reasonably precise estimates of stage-specific sensitivities.

MATERIALS AND METHODS

Data sources and searches

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were followed when conducting this meta-analysis. We searched PubMed, Web of Science, and Embase for original English-language human research articles from inception to June 14, 2019. The search of citing, cited and related articles was conducted using PubMed, Web of Science, Embase, and Google Scholar. Furthermore, we searched the Cochrane Library. Our search terms (see Supplement A, Supplementary Digital Content 1, <http://links.lww.com/AJG/B321>) covered expressions for FIT, CRC, diagnostic accuracy (sensitivity), and CRC staging. Those terms were agreed on after intense discussion, repeated sample searches, and comparison with articles retrieved by a recent previous meta-analysis of studies on diagnostic accuracy of FIT (2).

Study selection

To be eligible, studies had to report on the sensitivity and specificity of FIT for CRC detection. Sensitivities or sufficient information to calculate them had to be provided according to at least one distinct cancer stage (preferably stages I–IV separately). We preferably used the Union Internationale Contre le Cancer/American Joint Committee on Cancer (UICC/AJCC) staging system, which was the most commonly used system reported on. Alternatively, Duke's stages were also considered because they directly translate into a corresponding UICC/AJCC stage. Studies reporting only on T-stages were summarized separately because a wide range of combinations of T-, N-, and M-staging can result in the same UICC/AJCC stage.

Conduction of colonoscopy among all participants (not only FIT positives) was required for all studies. Three types of studies were identified as follows: (i) studies in which FIT was conducted in asymptomatic average-risk screening participants before colonoscopy (screening cohorts), (ii) studies which prospectively recruited participants who underwent colonoscopy for clarification of symptoms and not for primary screening (symptomatic/diagnostic cohorts), and (iii) retrospective studies in which patients diagnosed with CRC were included and compared with healthy controls (case-control). Studies focusing on high-risk populations (family history of CRC) were not considered because high-risk individuals are recommended to undergo colonoscopy

in shorter intervals rather than FIT screening (45), and among the previously meta-analyzed studies on accuracy of FIT among high-risk populations (46), no study reported on CRC stage-specific performance.

Data extraction

Two authors (T.N and Y.B.) extracted data independently. The following data were extracted from relevant articles: First author's last name, year of publication, study population characteristics such as country in which the study was conducted, study setting, number of participants and CRC cases, age ranges, sex distribution, FIT brand and cutoff, number of samples, and number of CRC cases detected and missed by FIT, stratified by the CRC stage. Initial disagreement in extracted data was resolved in consensus after further review and discussion. We contacted the corresponding authors of 33 articles identified from the previous systematic reviews and meta-analyses on FIT screening (2,47) and through cross-referencing to provide further information on stage-specific sensitivities of FIT in their studies. The corresponding authors of 3 articles (14,48,49) provided additional data on exact stage-specific case numbers and sensitivities. For 2 further articles in which relevant data were not directly reported (29,37), Hb concentrations for each CRC case (29) and data presented in a figure of the manuscript (37) were used to calculate the required number of true-positive and false-negative CRC cases. If the results were presented for several FIT cutoffs, we selected the cutoff closest to the cutoff recommended by the manufacturer or, if not applicable, a cutoff yielding a specificity comparable with the previously estimated overall specificity of FIT of 94% (50). If sensitivities were examined at different specificities, we again selected estimates for specificities closest to 94%. Hemoglobin-based FITs were preferably selected over hemoglobin-haptoglobin-based FITs if the results for both were selected because the former are much more commonly used.

Risk of bias and study quality assessment

We used the Quality Assessment of Diagnostic Accuracy Studies 2 instrument (QUADAS-2) (51) to assess the quality of included studies. Quality assessment was done by 2 authors (T.N. and Y.B.) in parallel, and disagreement was resolved in consensus after further discussion. Studies recruiting asymptomatic subjects for primary screening were categorized as low risk and case-control studies as high risk of bias in the category "Patient selection," and symptomatic/diagnostic cohorts were rated as unclear risk. A detailed description of the quality criteria and how they were adapted to our research question is given in Supplement B (see Supplementary Digital Content 2, <http://links.lww.com/AJG/B322>).

Potential publication bias was assessed using the methods of Deeks et al. (52) and Macaskill et al. (53).

Statistical analyses

Overall and stage-specific sensitivities of FIT were calculated as the number of CRC cases (total or of the respective stage from I to IV) with a positive FIT divided by the total or stage-specific number of CRCs. Specificities were calculated as the number of FIT-negative participants without CRC divided by the total number of participants without CRC. We used R (54) for all statistical computations. The R package "mada" (55) was used for computations of Clopper-Pearson CIs of sensitivities and specificities and for bivariate meta-analyses of sensitivities and

specificities using the Reitsma model (56). Meta-analysis was conducted only on studies reporting sensitivities for individual stages rather than stage groups because proportions of earlier and later stages may differ considerably between studies reporting combined stages, e.g., III and IV. For each stage, data from 2-by-2 tables were extracted and pooled using the function “reitsma” in the R package “mada,” which returns a pooled sensitivity and false-positive rate including 95% CIs. Pooled specificities were obtained from the same bivariate model but using all cases (irrespective of the stage) to calculate pooled sensitivities because sensitivity and specificity are always calculated together in the bivariate model.

Differences in pooled sensitivity between stages were assessed statistically using meta-regression techniques with stage as an explanatory variable and “stage I” as a reference group. Potential differences in pooled stage-specific sensitivity and specificity between settings were assessed using meta-regression with setting as an explanatory variable and “screening setting” as a reference group. Between-study heterogeneity was assessed by Higgins I^2

and P values of Cochrane’s Q for each meta-analysis. To investigate the influence of individual studies on the overall results, we conducted “leave-one-out” meta-analyses, excluding one study at a time and reported the resulting summary estimates of sensitivity and specificity and measures of residual heterogeneity. We conducted further sensitivity analyses using the results for the highest and lowest reported cutoffs for meta-analyses instead of the cutoffs recommended by the manufacturer in studies that reported on several cutoffs. Potential publication bias was assessed using the methods proposed by Deeks et al. (52), in which the diagnostic odds ratios as measures of overall effect size are plotted against the inverse of the square root of the effective sample size, and the method of Macaskill (53), which plots diagnostic odds ratios against the overall sample size.

RESULTS

Study selection process

A total of 13,013 unique articles were identified (Figure 1). After screening of titles and abstracts, 151 articles were given full-text

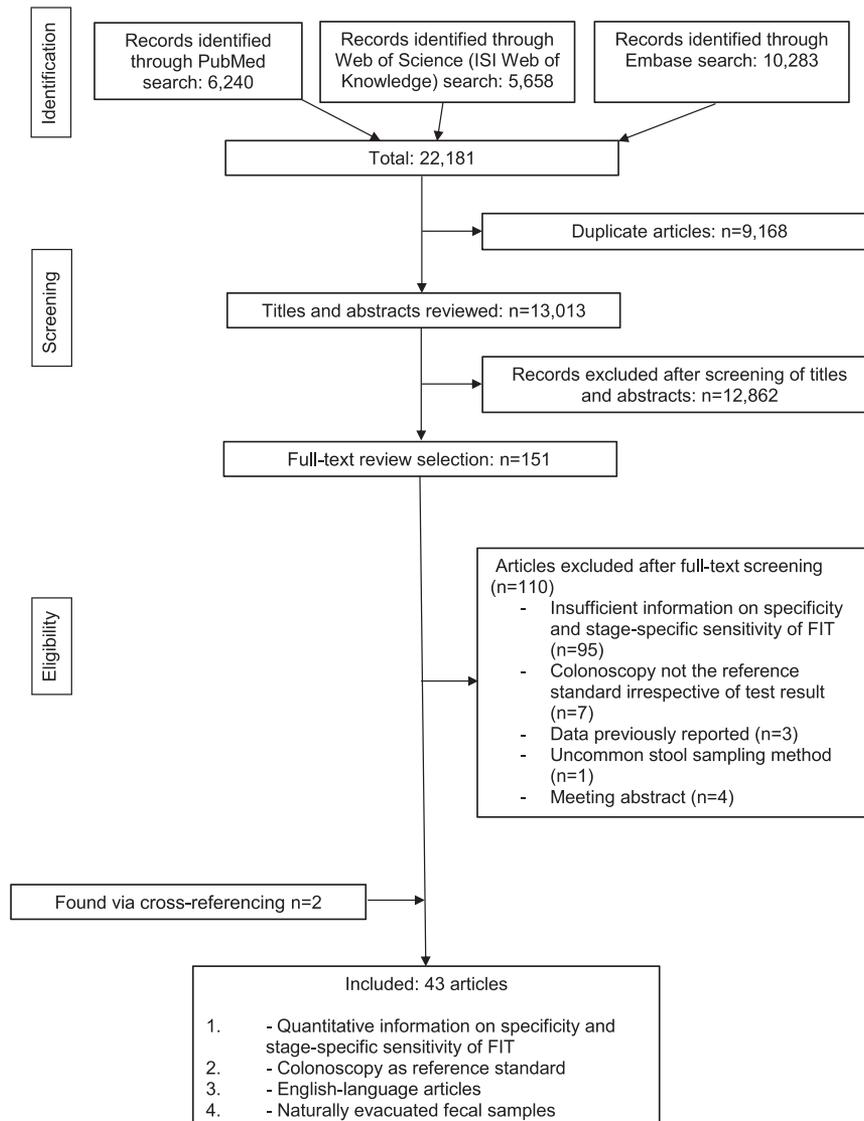


Figure 1. Process to identify articles for study.

assessment, and 110 of them were excluded. The most frequent exclusion criterion was lack of reporting on both the specificity and stage-specific sensitivity of FIT. Two relevant articles were found through cross-referencing (7,12). Forty-four studies (4–43,48,49,57,58) were in English, reported on the setting, specificity, and stage-specific sensitivity of FIT, and conducted colonoscopy in all participants, thus fulfilled the inclusion criteria.

Description of included studies

An overview of study characteristics is given in Table 1. We identified 12 screening cohorts (8,12,18,20,24,29,37,38,48,49,57,58) comprising 277 CRC cases, 18 symptomatic/diagnostic cohorts (4–7,10,11,13,16,17,19,21,22,25,27,28,34,35,40) comprising 869 CRC cases, and 14 case-control studies (9,14,15,23,26,30–33,36,39,41–43) with a total of 1,888 CRC cases. One report (23) also comprised screening participants but was classified as clinical (case-control) because approximately 80% of the CRC cases were recruited from a case-control study.

The most frequently reported FIT brands were FITs manufactured by Eiken Chemical (4,18,20,24,28,29,37,42,48,49,51,57,58). Other FITs used include Ridascreen Hemoglobin (16,23), FOB Gold (18,38), and InSure (11). Using stool from one sample was most common. Nine studies (4,7,10–12,18,40,41,49) reported the use of 2 samples. FIT cutoffs ranged from 2 to 67 $\mu\text{g Hb/g stool}$, but most studies reported thresholds between 10 and 20 $\mu\text{g/g}$. Ten studies (4,6,7,11,14–17,32,41) used a qualitative FIT, and 2 studies (6,10) used a quantitative FIT but did not report on its positivity threshold.

Most study reports were published between 2003 and 2017. Three older studies comprising clinically detected CRC cases published between 1995 and 1998 (4,5,41) were also included. Studies were conducted among participants from Japan (4,8,12,13,26,40–42,57), Germany (5,16–18,23,38), Taiwan (24,27,34,58), Australia (7,11,35), the United States (29,30,33), the Netherlands (15,19,22), China (10,39,43), Spain (28,48,49), Thailand (14,37), Hong Kong (6,36), South Korea (9,20), the United Kingdom (25), Italy (21), and Saudi Arabia (32).

Mean or median age ranged from 48.2 to 67 years was reported. Shares of men ranged mostly between ~40% and 65%, except for 3 studies (4,8,12) with a predominantly male study population (72%–90%).

Results of the study quality assessment

We summarized the results of the QUADAS-2 study quality assessment in Table 2. Owing to exclusion of studies that did not use colonoscopy as a reference standard, all included studies were deemed as low risk of bias in the category “reference standard.” However, not all studies described whether CRC staging was conducted by pathologists, which we required for a “low risk” evaluation in the applicability concerns for the reference standard. Overall, studies were of moderate-to-high quality.

Meta-analysis of stage- and setting-specific performance of FIT

Distributions of CRC stages across studies are presented in Table 3. Screening cohorts included higher shares of early-stage CRCs (70% stage I or II) than symptomatic/diagnostic cohorts (57%) and case-control studies (54%).

Pooled specificity was 87% in screening cohorts, 87% in symptomatic/diagnostic cohorts, and 93% in case-control studies (Table 4). In Table 4, we also show individual study

and overall stage-specific sensitivities and specificities of FIT, grouped by setting in which the study was conducted. In most individual studies, sensitivities were higher in more advanced CRC stages (II–IV) compared with stage-I-CRCs. Pooled estimates (95% CIs) of stage-specific sensitivities were similar across the 3 different study settings, with widely overlapping CIs (which were though much narrower for symptomatic/diagnostic cohorts and case-control studies due to the much higher number of CRC cases in these groups than in the screening cohorts). Pooled estimates across the study settings were 80% (95% CI 74–84%), 82% (95% CI 77–87%), and 79% (95% CI 70–86%) for stages II, III, and IV, respectively, compared with 73% (95% CI 65–79%) for stage I (*P* values for differences to stage I were 0.02, 0.005, and 0.28, respectively). Studies reporting on FIT sensitivity by T-stages showed substantially higher sensitivities in stages T2, T3, and T4 (79%, 83%, and 66%) than in T1 CRCs (40%). *P* values for differences to T1 were 0.02, 0.004, and 0.009, respectively.

Sensitivity analyses

Leaving out one study at a time from each meta-analysis had only minor influence on pooled sensitivities and specificities (see Table, Supplementary Digital Content 3, <http://links.lww.com/AJG/B319>). Using the highest or lowest reported cutoff instead of the cutoff recommended by the manufacturer likewise did not change the summary estimates of sensitivity or specificity materially. Point estimates of pooled sensitivity using the highest reported cutoffs were 71%, 79%, 82%, and 79% for stages I, II, III, and IV, respectively. Using the lowest reported cutoffs, the respective estimates were 72%, 79%, 81%, and 80%.

Assessment of heterogeneity

Statistical heterogeneity was small. *P* values of Cochran's *Q* were >0.2 in all analyses. Furthermore, within each cancer stage, sensitivities did not differ significantly by the setting (all $P \geq 0.27$, see Table, Supplementary Digital Content 4, <http://links.lww.com/AJG/B320>).

Assessment of publication bias

Plots of effect size vs effective or actual sample size showed no trend, giving no indication for publication bias being present (see Figure, Supplementary Digital Content 5, <http://links.lww.com/AJG/B318>).

DISCUSSION

We conducted a systematic review of studies reporting the specificity and stage-specific sensitivity of FITs for CRC detection. Meta-analyses of suitable studies suggested that the sensitivity of FIT was considerably lower, by approximately 8% units, in CRC stage I compared with stages II and IV. Meta-regression suggested that sensitivity for stage I cancers was significantly lower than sensitivity for stages II + III. In numbers, 1 of 4 stage-I-CRCs was missed by FIT, compared with less than 1 of 5 CRCs at later stages. A stronger gradient in sensitivity was seen for lower vs higher T-stages, although the T-stage-specific results were reported from a few studies only. Estimated sensitivity for T1 CRCs was 40%, thus only slightly higher than the previously estimated sensitivity of FIT for advanced adenomas (16%–34%) (59). By contrast, stage-specific sensitivities did not differ significantly across different study settings.

Table 1. Characteristics of the studies investigating the stage-specific sensitivity and specificity of FIT in a clinical or screening setting

First author (reference)	Year	Country	N total	N CRC	Mean or median age	Age range	Male (%)	FIT brand	FIT cutoff (μg/g)	No. of FIT samples
Screening cohorts										
Morikawa et al. (8)	2005	Japan	21,805	79	48.2	20–91	72.0	Magstream 1000/Hem SP	67	1
Nakazato et al. (12)	2006	Japan	3,090	19	52.7	NR	84.8	OC-Hemodia	16	2
Graser et al. (18)	2009	Germany	285	1	60.5	>50	55.0	FOB Gold	2.4	2
Parra-Blanco et al. (48)	2010	Spain	2,280	14	62.7	50–79	32.8	OC-Light	10	1
Park et al. (20)	2010	South Korea	770	13	59.3	NR	51.4	OC-Sensa Micro	20	1
Chiu et al. (24)	2013	Taiwan	18,296	28	59.8	>50	59.2	OC-Light	10	1
Hernandez et al. (49)	2014	Spain	779	5	57.8	50–69	49.6	OC-Sensor	10/15/20/23/30/40	1 (2) ^a
Imperiale et al. (29)	2014	USA/ Canada	9,989	65	NR	50–84	46.3	OC FIT-CHEK	20	1
Wakamura et al. (57)	2015	Japan	919	12	57.9	NR	61.4	OC-Hemodia + OC-Hemocatch S	30/10 ^c	1
Aniwan et al. (37)	2017	Thailand	1,479	9	60.4	50–75	38.3	OC-Sensor Diana	5/10/30/40	1
Brenner et al. (38)	2017	Germany	3,466	29	NR	50–79	49.7	FOB Gold	17	1
Chang et al. (58)	2017	Taiwan	6,198	3	59.0	>50	51.1	OC-Sensor	20	1
Subtotal number of study participants			69,356	277	—	—	—	—	—	—
Symptomatic/diagnostic cohorts										
Yoshinaga et al. (4)	1995	Japan	885	23	53.4	33–82	89.9	OC-Hemodia	40	2
Sieg et al. (5)	1998	Germany	739	43	NR	21–85	52.8	DAKO	2	1
Wong et al. (6)	2003	Hong Kong	135	9	58	38–90	42.0	FlexSure OBT	NR	1
Young et al. (7)	2003	Australia	443	36	NR	24–90	51.8	InSure [FlexSure OBT]	NA	2
Li et al. (10)	2006	China	324	49	53.5	18–68	57.6	NR [Wanhua-Puman Biol.]	NR	2
Smith et al. (11)	2006	Australia	161	7	66.2	NR	NR	InSure	NA	2
Hirata et al. (13)	2007	Japan	872	36	NR	36–81	NR	NA	100 ng/mL	1
Karl et al. (16)	2008	Germany	551	101	65.1	NR	45.8	Ridascreen Hemoglobin	NA	1
Shastri et al. (17)	2008	Germany	640	55	NR	42–88	41.4	Prevent ID	NA	1
Oono et al. (40)	2010	Japan	1,073	91	65.0	26–89	52.4	Nescauto Hemo Plus	100 ng/mL	2
Oort et al. (19)	2010	Netherlands	1,821	59	59.6	18–86	43.1	OC-Sensor	100 ng/mL	1
Terhaar sive Droste et al. (22)	2011	Netherlands	2,145	79	61.8	40–89	46.2	OC-Sensor Micro	10/15/.../40	1
Kalimutho et al. (21)	2011	Italy	192	11	61	19–82	44.8	NR [MP Biomedical LLC]	50 ng/mL	1
Kaul et al. (25)	2013	UK	112	17	66	34–87	45.5	OC-Light	40 ng/mL	1
Lee et al. (27)	2013	Taiwan	3,172	39	53.0	19–92	60.5	OC-Sensor	20	1
Cubiella et al. (28)	2014	Spain	787	97	67.0	22–91	51.1	OC-Sensor	20	1
Chen et al. (34)	2016	Taiwan	60	51	65.8	NR	58.3	NR (Taipower)	200 ng/mL	1
Symonds et al. (35)	2016	Australia	1,381	66	64.1	41–85	49.4	OC-Sensor	60	1
Subtotal number of study participants			15,493	869	—	—	—	—	—	—

Table 1. (continued)

First author (reference)	Year	Country	N total	N CRC	Mean or median age	Age range	Male (%)	FIT brand	FIT cutoff (µg/g)	No. of FIT samples
Case-control studies										
Nakama et al. (41)	1997	Japan	552	184	NR	NR	NR	NA	NA	1/2/3 ^b
Sohn et al. (9)	2005	South Korea	3,794	316	48.9	15–78	56.7	OC-Hemodia	20	1
Lohsiriwat et al. (14)	2007	Thailand	164	100	63	23–82	49.0	OC-Light	NA	1
Mulder et al. (15)	2007	Netherlands	162	52	NR	27–91	51.4	OC-Light + Immocare	NA	1
Tao et al. (23)	2012	Germany	597	66	NR	NR	54.8	Ridascreen Hemoglobin	2	1
Koga et al. (26)	2013	Japan	224	117	NR	30–84	59.0	OC-Hemocatch	50 ng/mL	1
Johnson et al. (30)	2014	USA	290	101	NR	50–84	38.7	OC FIT-CHEK	100 ng/mL	NR
Kim et al. (dev. set) (31)	2014	Korea	132	81	58.2	NR	NR	OC-Sensor	100 ng/mL	NR
Kim (validation set) et al. (31)	2014	Korea	194	94	56.0	NR	NR	OC-Sensor	100 ng/mL	NR
Elsafi et al. (32)	2015	Saudi Arabia	277	23	63.8	29–89	NR	RAPEPKT313	50 ng/mL	1
Baxter et al. (33)	2016	USA	172	120	60	29–89	NR	OC FIT-CHEK	100 ng/mL	1
Wong et al. 2017 (36)	2017	Hong Kong	309	104	61.8	NR	64.7	OC-Sensor Micro	20	1
Xie et al. (39)	2017	China	345	424	NR	>40	NR	Gold gel stripe	200 ng/mL	1
Suehiro et al. (42)	2018	Japan	138	18	67	36–91	67.3	OC-Hemodia	20	NR
Xie et al. (43)	2018	China	248	88	66	NR	56.0	WWT/FA160-auto analyzer	200 ng/mL	NR
Subtotal number of study participants			7,598	1,888	—	—	—	—	—	—
Total number of study participants (any)			92,447	3,034	—	—	—	—	—	—
CRC, colorectal cancer; FIT, fecal immunochemical test; NA, not applicable; NR, not reported.										
^a Hernandez et al. reported overall sensitivities and specificities for both one- and two-sample FIT, but the results of one-sample FIT were used for meta-analysis.										
^b Nakama et al. reported the results for 1-, 2-, and 3-sample FIT. For our analyses, we used the results from 2-sample FIT because 2 samples resulted in a “typical” specificity of FIT (92.5%), unlike one-sample FIT (97.6%), and three-sample FITs are uncommon in screening practice.										
^c OC-Hemodia with a cutoff of 150 ng/mL was used until March 2008, and OC-Hemocatch is used since April 2008 with a 50 ng/mL cutoff (personal correspondence with the author).										

A gradient in FIT sensitivity by stage is plausible because later-stage CRCs are typically larger and bleed stronger. An even stronger gradient in sensitivity across T-stages is likewise plausible because the size of the primary tumor is directly related to intestinal bleeding, unlike nodal involvement (N) or metastases (M). However, these suggested associations between T or overall TNM stage and the sensitivity of FIT have not been quantified to date, and investigations of potential setting-specific differences in FIT sensitivity by stage were lacking. The lack of a further increase in sensitivity for stage IV compared with stage III might result from iron-deficiency anemia in more advanced tumors (60). Early-stage CRCs (stages I and II) are associated with considerably higher survival rates than stages III and IV (61). Thus, the sensitivity of FIT and other screening tests for these early stages would be much more relevant in clinical practice than sensitivity for late-stage CRCs.

Comparable stage-specific sensitivities across different study types have been suggested previously (44) and allowed for joint consideration of all studies reporting on FIT sensitivity by the CRC stage. CRC case numbers increased 11-fold compared to an

analysis based on screening cohorts alone, resulting in much more precise estimates of stage-specific sensitivities of FIT. In particular, the width of CIs for the sensitivity of stage I, II, III, and IV CRC decreased from 32 to 14, from 24 to 10, from 29 to 10, and from 53 to 16 percentage points, respectively. The specificity of FIT was highest in case-control studies. Studies of this type often include healthy subjects as controls who are also typically less prone to potential alternative sources of bleeding (e.g., hemorrhoidal bleeding), making false-positive results unlikely. A similar specificity was observed in symptomatic/diagnostic cohorts and in screening cohorts. In the former, non-neoplastic gastrointestinal disorders may cause bleeding and, thus, reduce the specificity of FIT for colorectal neoplasia detection. In screening cohorts, the potentially older mean age of participants may contribute to a reduced specificity compared with symptomatic/diagnostic cohorts (62).

A meta-analysis on the overall sensitivity and specificity of FIT for CRC detection in screening settings irrespective of the stage was recently published (2). Reported overall levels of the sensitivity (77%–94%) and specificity (85%–96%) of FIT that are

Table 2. Quality assessment of diagnostic accuracy studies 2 instrument risk of bias assessment

Study (first author, year [ref.])	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Yoshinaga et al. (4)	😊	😊	😊	😊	?	😊	😊
Nakama et al. (41)	😞	?	😊	😊	😞	😞	?
Sieg et al. (5)	😊	?	😊	😊	?	😞	😊
Wong et al. (6)	😊	😊	😊	😊	?	😊	😊
Young et al. (7)	😊	?	😊	😞	?	😊	😊
Morikawa et al. (8)	😊	😊	😊	😊	😊	😊	😊
Sohn et al. (9)	😞	😊	😊	😊	😞	😊	😊
Li et al. (10)	😊	?	😊	😊	?	😊	😊
Nakazato et al. (12)	😊	?	😊	😊	😊	😞	😊
Smith et al. (11)	😊	😊	😊	😊	?	😊	😊
Hirata et al. (13)	😊	😞	😊	😊	?	😞	?
Lohsiriwat et al. (14)	😞	?	😊	😊	😞	😊	😊
Mulder et al. (15)	😞	😊	😊	😞	😞	😊	😊
Karl et al. (16)	😊	😊	😊	😊	?	😊	😊
Shastri et al. (17)	😊	😊	😊	😊	?	😊	?
Graser et al. (18)	😊	😊	😊	😊	😊	😊	😊
Oono et al. (40)	😊	?	😊	😊	?	😊	?
Oort et al. (19)	😊	😊	😊	😞	?	😊	😊
Park et al. (20)	😊	?	😊	😊	😊	😊	?
Parra-Blanco et al. (48)	😊	?	😊	😞	😊	😊	?
Kalimutho et al. (21)	😊	?	😊	😊	?	😊	😊
Terhaar sive Droste et al. (22)	😊	😊	😊	😊	?	😊	😊
Tao et al. (23)	? ^a	😊	😊	😊	?	😊	😊
Chiu et al. (24)	😊	😊	😊	😊	😊	😊	😊
Kaul et al. (25)	😊	?	😊	😊	?	😊	😊
Koga et al. (26)	😞	?	😊	😊	😞	😊	?
Lee et al. (27)	😊	?	😊	😊	?	😊	😊
Cubiella et al. (28)	😊	😊	😊	😊	?	😊	?

Table 2. (continued)

Study (first author, year [ref.])	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Hernandez et al. (49)	😊	😊	😊	😊	😊	😊	?
Imperiale et al. (29)	😊	😊	😊	😊	😊	😊	😊
Johnson et al. (30)	😞	?	😊	😊	😞	😊	?
Kim et al. (31)	😞	?	😊	😊	😞	😊	😊
Elsafi et al. (32)	😞	?	😊	😊	😞	😊	😊
Wakamura et al. (57)	😊	?	😊	😞	😊	?	?
Baxter et al. (33)	😞	?	😊	😊	😞	😊	?
Wong et al. (36)	😞	😊	😊	😊	😞	😊	😊
Chen et al. (34)	😞	?	😊	😞	?	😊	😊
Symonds et al. (35)	😊	😊	😊	😞	?	😊	?
Aniwan et al. (37)	😊	😊	😊	😊	😊	😊	?
Brenner et al. (38)	😊	😊	😊	😞	😊	😊	😊
Chang et al. (58)	😊	😊	😊	😊	😊	😊	😊
Xie et al. (39)	😞	😊	😊	😊	😞	😊	😊
Suehiro et al. (42)	😞	?	😊	😊	😞	😊	?
Xie et al. (43)	😞	?	😊	😞	😞	?	😊

😊 Low risk; 😞 high risk; and ? unclear risk.

^aStudy population comprised approximately 4/5 of participants recruited from a clinical setting and 1/5 of participants from a screening setting.

in the same range as ours (screening cohorts: 84% and 87%, respectively; all studies: 81% and 89%, respectively), with some variation according to FIT cutoffs. Unlike their meta-analysis, we did not restrict our search to studies conducted in a screening setting which substantially increased the number of cases that could be included and enabled performance of stage-specific analyses, an additional inclusion criterion in our meta-analysis. In 2017, Katsoula et al. (46) published a meta-analysis of FIT accuracy among high-risk individuals with a family history of CRC. Including only studies that used colonoscopy as a reference standard, they found a high sensitivity and specificity of FIT for CRC detection (93% and 91%, respectively). The high summary estimate of sensitivity seems to suggest that FIT might perform better in high-risk subjects compared with the general population. However, owing to the inclusion of only 40 CRC cases overall, the 95% CI for sensitivity was very wide (53%–99%), and the results were thus also compatible with a similar or even lower sensitivity than derived for CRC cases irrespective of family history in our meta-analysis.

It has been suggested that FIT has lower miss rates for distal CRC than for proximal colon cancer (59). Because no study

reported on the sensitivity of FIT stratified by both the stage and location, we could not investigate the joint role of anatomic location and stage for FIT performance. Some site-specific variation in sensitivity by stage is conceivable. For example, early-stage CRCs with minimal bleeding might have a larger chance to be detected when located in the distal colon or rectum rather than the proximal colon, whereas site differences may be less relevant for advanced stages with more extensive bleeding.

To the best of our knowledge, this is the first systematic review and meta-analysis of the stage-specific (rather than overall) sensitivity of FIT for CRC detection. A particular strength is the comprehensive search in 4 databases. In total, data from 44 studies comprising >92,000 participants, thereof ~3,000 CRC cases, were used. We did not restrict our search to a single geographical region, thereby achieving increased external validity. Data were extracted by 2 authors independently. Bivariate meta-analysis was used to jointly estimate pooled sensitivities and specificities, taking into account the cutoff dependence of and negative correlation between the 2 measures. We adhered to the standards of reporting for systematic reviews and meta-analyses

Table 3. CRC stage distribution of included studies

First author, year (reference)	0/I/A N (%)	II/B N (%)	III/C N (%)	IV/D N (%)	Early stage ^a N (%)	With stage N
Screening cohorts						
Morikawa et al. (8)	36 (52)	10 (14)	23 (33)		46 (66)	69
Graser et al. (18)	0 (0)	0 (0)	1 (100)	0 (0)	0 (0)	1
Park et al. (20)		10 (77)	3 (23)		10 (77)	13
Hernandez et al. (49)	3 (60)	1 (20)	1 (20)	0 (0)	4 (80)	5
Imperiale et al. (29)	29 (45)	21 (32)	10 (15)	4 (6)	50 (77)	64
Wakamura et al. (57)	6 (60)	3 (30)	1 (10)	0 (0)	9 (90)	10
Aniwan et al. (37)	0 (0)	7 (78)	2 (22)	0 (0)	7 (78)	9
Brenner et al. (38)	10 (34)	4 (14)	12 (41)	3 (10)	14 (48)	29
Chang et al. (58)	1 (33)	1 (33)	1 (33)	0 (0)	2 (67)	3
Total					142 (70)	203
Symptomatic/diagnostic cohorts						
Yoshinaga et al. (4)	10 (43)	8 (35)	3 (13)	2 (9)	18 (78)	23
Sieg et al. (5)	16 (37)	4 (9)	16 (37)	7 (16)	20 (47)	43
Wong et al. (6)	1 (11)	4 (44)	1 (11)	3 (33)	5 (56)	9
Young et al. (7)		21 (58)	15 (42)		21 (58)	36
Li et al. (10)	5 (10)	15 (31)	16 (33)	4 (8)	20 (50)	40
Smith et al. (11)	3 (43)	1 (14)	3 (43)	0 (0)	4 (57)	7
Karl et al. (16)	23 (27)	27 (32)	12 (14)	23 (27)	50 (59)	85
Shastri et al. (17)		36 (65)	19 (35)		36 (65)	55
Oono et al. (40)	70 (77)		21 (23)		≥70 (49)	91
Oort et al. (19)		28 (47)	31 (53)		28 (47)	59
Terhaar sive Droste et al. (22)		38 (51)	36 (49)		38 (51)	74
Kalimutho et al. (21)	2 (18)	6 (55)	3 (27)	0 (0)	8 (73)	11
Kaul et al.(25)	1 (7)	5 (36)	5 (36)	3 (21)	6 (43)	14
Lee et al. (27)		21 (54)	18 (46)		21 (54)	39
Cubiella et al. (28)	20 (21)	23 (24)	40 (41)	13 (13)	43 (44)	96
Chen et al. (34)		23 (45)	28 (55)		23 (45)	51
Symonds et al. (35)	17 (26)	25 (38)	17 (26)	7 (27)	42 (64)	66
Total					453 (57)	799
Case-control studies						
Nakama et al. (41)	86 (47)	45 (24)	53 (29)	0 (0)	131 (71)	184
Sohn et al. (9)	41 (13)	92 (29)	183 (58)	0 (0)	133 (42)	316
Lohsiriwat et al. (14)	14 (14)	25 (25)	61 (61)		39 (39)	100
Tao et al. (23)	21 (32)	16 (24)	29 (44)		37 (56)	66
Koga et al. (26)		76 (65)	41 (35)		76 (65)	117
Johnson et al. (30)	26 (32)	20 (25)	23 (28)	12 (15)	46 (57)	81
Elsafi et al. (32)	2 (9)	15 (65)	3 (13)	3 (13)	17 (64)	23
Baxter et al. (33)	39 (33)	35 (29)	32 (27)	10 (8)	74 (62)	116
Suehiro et al. (42)	14 (78)		4 (22)		≥14 (78)	18
Xie et al. (43)	5 (5)	36 (35)	58 (56)	4 (4)	41 (40)	103
Wong et al. (36)	22 (21)	31 (30)	39 (38)	12 (12)	53 (51)	104
Total					661 (54)	1,228
Total all studies					1,256	2,230

Table 3. (continued)

First author, year (reference)	T1 N (%)	T2 N (%)	T3 N (%)	T4 N (%)	Early stage ^a N (%)	With stage N
T-staging ^b						
Sohn et al. (9)	13 (4)	40 (13)	191 (60)	72 (23)	NA	316
Nakazato et al. (12)		<i>18 (95)</i>		<i>1 (5)</i>	NA	19
Hirata et al. (13)	17 (47)		<i>19 (53)</i>		NA	36
Mulder et al. (15)		<i>9 (19)</i>		<i>38 (81)</i>	NA	47
Parra-Blanco et al. (48)	10 (71)	3 (21)	0 (0)	1 (7)	NA	14
Chiu et al. (24)	18 (78)		<i>5 (22)</i>		NA	23
Kim et al. (31)	14 (17)	10 (12)	47 (58)	10 (12)	NA	81
Wakamura et al. (57)	7 (58)	0 (0)	4 (33)	1 (8)	NA	12
Xie et al. (39)	103 (24)		<i>321 (76)</i>		NA	424
Aniwan et al. (37)	0 (0)	0 (0)	8 (89)	1 (11)	NA	9
Chang et al. (58)	0 (0)	1 (33)	2 (67)	0 (0)	NA	3
Total					—	984

CRC, colorectal cancer; NA, not applicable.
 Estimates in italics: not included in meta-analysis.
^aAJCC/UICC stages 0/I or II, Dukes A or B.
^bStudies reporting sensitivities according to T-stages were conducted in a case-control setting (clinically recruited), except for Chiu et al., Nakazato et al., and Wakamura et al. (screening setting) and 1 symptomatic/diagnostic cohort (Hirata et al.).

(PRISMA) and for studies on diagnostic accuracy (STARD). Study quality was assessed using the QUADAS-2 tool specifically designed for studies of diagnostic accuracy. By focusing on studies using colonoscopy as a reference standard, verification bias which may cause a systematic overestimation of sensitivity and underestimation of specificity was ruled out.

Several limitations of our meta-analysis should be kept in mind. Despite extensive search in 4 databases and cross-referencing, it cannot be ruled out that we missed a relevant article. Language, publication, and reporting bias are conceivable. Although no statistical hypothesis was tested in the underlying studies, pronounced stage-specific differences could be more likely to be reported than small differences. We included studies irrespective of the FIT used, despite differences in FIT types (quantitative/qualitative), brands, and cutoffs. It has been suggested, however, that different quantitative FITs show similar performance characteristics when adjusting the cutoffs to yield equal specificities (63). More studies would be needed for subgroup analyses, e.g., according to the FIT brand, country of study conduct, and characteristics of study populations. Although no differences were observed in stage-specific sensitivities between different study types, the results of this meta-analysis should be interpreted with caution because of pooling of data from colonoscopies with various indications. Accuracy estimates may also be influenced by disease prevalence (64) and severity, which varied between the 3 study types. However, differences in prevalence would introduce spectrum bias only in case of an imperfect reference standard (65). Disease severity was addressed by stratifying according to the CRC stage. Somewhat uneven distributions within each stage (regarding substages, e.g., IIA or IIB) might remain, but those differences would likely be small and random. Another limitation of our meta-analysis is that few studies reported how they handled patients undergoing

postpolypectomy surveillance regarding inclusion or exclusion. Among the 3 studies reporting on that criterion, one study (12) excluded such patients, whereas the other 2 studies (19,22) included them (but did not stratify the results according to this criterion). Thus, although desirable, it is unfortunately not feasible from the given data to conduct a subgroup analysis comparing patients from this group with others.

Our study suggests that CRC stage distribution is a major determinant of diagnostic performance of FIT. Because staging is routinely conducted, studies should report FIT results according to CRC stages. Ideally, complete TNM stages would be reported (from which AJCC/Duke's stages can be derived), given that our study also suggests that T-stage is a particularly important determinant of FIT sensitivity. Still, the findings of this study require stringent validation. Such validation could take place through clinics and practices in which FITs are conducted for screening participants and symptomatic individuals and could be performed also among newly diagnosed CRC patients. By following standard operating procedures, potential variation due to different FITs, cutoffs, buffer solutions, delay until Hb measurement, etc. could be reduced to a minimum. Future studies should also consider reporting results on several FIT cutoffs, which would facilitate pooling of studies by using cutoffs yielding similar specificity. Besides, reporting accuracy estimates or ideally exact Hb concentrations for each CRC case (e.g. Aniwan et al. (37)) would enable detailed analyses of sensitivity/specificity for a range of possible cutoffs, along with its implications for number of colonoscopies needed to detect one CRC (or precursors) (66). Those estimates would be particularly useful for countries with limited colonoscopy resources and allow for identifying an optimal FIT positivity threshold depending on characteristics such as CRC/adenoma prevalence, available colonoscopy resources, and costs associated with colonoscopy or other screening modalities.

Table 4. Summary of the results on FIT performance of included studies for CRC, stratified by the stage

First author (reference)	Sensitivity % (95% CI)				Specificity ^a % (95% CI)
	I/A	II/B	III/C	IV/D	
Screening cohorts					
Morikawa et al. (8)	53 (36–70)	70 (35–93)	78 (56–93)		95 (94–95)
Graser et al. (18)	—	—	100 (3–100)	—	84 (79–88)
Parra-Blanco et al. (48)	100 (69–100)	100 (29–100)	—	100 (3–100)	95 (92–97)
Park et al. (20)	80 (44–97)		100 (29–100)		90 (88–92)
Hernandez et al. (49)	100 (29–100)	100 (3–100)	100 (3–100)	—	94 (92–96)
Imperiale et al. (29)	66 (46–82)	76 (53–92)	90 (55–100)	75 (19–99)	93 (93–94)
Wakamura et al. (57)	100 (54–100)	67 (9–99)	100 (3–100)	—	72 (69–75)
Aniwan et al. (100 ng/mL) (37)	—	100 (59–100)	100 (16–100)	—	93 (92–95)
Chang et al. (58)	100 (3–100)	100 (3–100)	100 (3–100)	—	95 (95–96)
Brenner et al. (38)	90 (55–100)	100 (40–100)	100 (74–100)	100 (29–100)	90 (89–91)
Summary estimate (indiv. stages)	75 (56–88)	77 (63–87)	85 (65–94)	79 (42–95)	87 (75–94)
Symptomatic/diagnostic cohorts					
Yoshinaga et al. (4)	90 (60–98)	100 (63–100)	100 (29–100)	100 (16–100)	84 (81–86)
Sieg et al. (5)	94 (70–100)	75 (19–99)	100 (79–100)	100 (59–100)	80 (77–83)
Wong et al. (6)	0 (0–98)	100 (40–100)	100 (3–100)	100 (29–100)	91 (85–96)
Young InSure [FlexSure OBT] (7)	76 (53–92)		73 (45–92) [87 (60–98) ^b]		90 (87–93)
Li et al. (10)	40 (5–85)	93 (68–100)	94 (70–100)	100 (40–100)	83 (78–87)
Smith et al. (11)	100 (29–100)	100 (3–100)	100 (29–100)	—	81 (73–87)
Karl et al. (hemoglobin, sp. 95%) (16)	74 (52–90)	85 (66–96)	92 (62–100)	83 (61–95)	90 (87–93)
Shastri et al. (17)	64 (46–79)		84 (60–97)		93 (91–95)
Oono et al. (40)	90 (68–99)		95 (76–100)		90 (87–92)
Oort et al. (19)	75 (55–89)		97 (83–100)		91 (90–92)
Terhaar sive Droste et al. (100 ng/mL) (22)	82 (66–92)		97 (85–100)		90 (89–91)
Kalimutho et al. (21)	100 (16–100)	100 (54–100)	100 (29–100)	—	80 (73–86)
Kaul et al. (25)	100 (3–100)	100 (48–100)	100 (48–100)	100 (29–100)	86 (78–93)
Lee et al. (27)	81 (58–95)		83 (59–96)		97 (96–97)
Cubiella et al. (28)	80 (52–96)	91 (72–99)	88 (73–96)	92 (64–100)	77 (74–81)
Chen et al. (34)	70 (47–87)		96 (82–100)		56 (22–86)
Symonds et al. (60 µg/g) (35)	59 (33–82)	68 (46–85)	59 (33–82)	71 (29–96)	92 (90–93)
Summary estimate (indiv. stages)	79 (68–86)	88 (80–93)	85 (75–91)	87 (76–93)	87 (83–90)
Case-control studies					
Nakama et al. (41) (2-day method)	85 (76–92)	91 (79–98)	91 (79–97)	—	96 (93–98)
Sohn et al. (9)	63 (47–78)	79 (70–87)	79 (72–84)	—	99 (98–99)
Lohsiriwat et al. (14)	71 (42–92)	88 (69–97)	—	—	94 (85–98)
Tao et al. (23)	33 (15–57)	75 (48–93)	83 (64–94)		90 (87–93)
Koga et al. (26)	58 (46–69)		66 (49–80)		98 (93–100)
Johnson et al. (30)	65 (44–83)	80 (56–94)	83 (61–95)	58 (28–85)	97 (94–99)
Elsafi et al. (32)	50 (1–99)	60 (32–84)	67 (9–99)	100 (29–100)	90 (86–94)
Baxter et al. (33)	82 (66–92)	91 (77–98)	89 (74–97)	90 (55–100)	90 (87–93)
Wong et al. (36)	68 (45–86)	74 (55–88)	77 (61–89)	67 (35–90)	91 (87–95)
Suehiro et al. (42)	29 (8–58)		100 (40–100)		90 (78–97)
Xie et al. (43)	80 (28–99)	56 (38–72)	69 (55–80)	25 (6–81)	70 (59–80)

Table 4. (continued)

First author (reference)	Sensitivity % (95% CI)				Specificity ^a % (95% CI)
	I/A	II/B	III/C	IV/D	
Summary estimate (indiv. stages)	64 (50–76)	78 (69–85)	78 (73–82)	64 (46–78)	93 (88–96)
All studies, summary estimate	73 (65–79)	80 (74–84)	82 (77–87)	79 (70–86)	89 (85–92)
First author (reference)	Sensitivity of FIT by the CRC stage (T-stage of TNM classification)				Specificity ^a % (95% CI)
	T1	T2	T3	T4	
Studies reporting T-stages ^e					
Sohn et al. (9)	38 (14–68)	75 (59–87)	79 (73–85)	79 (68–88)	99 (98–99)
Nakazato et al. (12), ^e	50 (26–74)		100 (3–100)		87 (86–88)
Hirata et al. (13)	12 (1–36)		84 (60–97)		90 (88–92)
Mulder et al. (15) (OC-Light)	100 (66–100)		92 (79–98)	—	93 (86–98)
Mulder et al. (15) (Immocare)	100 (66–100)		89 (75–97)	—	93 (86–98)
Parra-Blanco et al. (48)	100 (16–100)	100 (63–100)	100 (29–100)	—	95 (92–97)
Chiu et al. (24), ^e	67 (41–87) ^c		100 (69–100)		93 (93–93)
Kim et al. (30), ^d	13 (3–33)	44 (21–69)	59 (49–69)	74 (54–89)	99 (96–100)
Wakamura et al. (57)	100 (59–100)	—	100 (16–100)	0 (0–98)	71 (68–74)
Xie et al. (39)	39 (29–49)	—	56 (49–62)	—	91 (86–94)
Aniwan et al. (100 ng/mL) (37), ^e	—	—	100 (63–100)	100 (3–100)	93 (92–95)
Chang et al. (58)	—	100 (3–100)	100 (16–100)	—	95 (95–96)
Summary estimate (indiv. stages)	40 (21–64)	79 (51–93)	83 (68–91)	66 (45–82)	93 (89–96)

95% CI, 95% confidence interval; CRC, colorectal cancer; FIT, fecal immunochemical test; sp., specificity.
^aSpecificity was defined by subjects free of advanced neoplasms (CRC or advanced adenoma).
^bSensitivity when using the results for FlexSure OBT instead of InSure.
^cIncluding *in situ* carcinomas.
^dSensitivities were reported separately for a development and validation set of a marker combination but combined for the present analyses.
^eChiu et al., Nakazato et al., Aniwan et al., and Parra-Blanco et al. were conducted in a screening setting, and the other studies reporting sensitivities according to T-stages were conducted in a case-control setting (clinically recruited).

This meta-analysis focused on the sensitivity and specificity of FIT in a single screening round. However, actual FIT screening is conducted in consecutive rounds, typically annually or biennially. Although adenoma detection rates would decrease in subsequent screening rounds because of removal of adenomas in FIT positives, and screening is associated with a shift to earlier cancer stages (67), it is unlikely that repeated screening would influence stage-specific sensitivity for CRC. Although repeated screening was not investigated in this study, our results may provide valuable input for microsimulation models of repeat screening scenarios, which crucially depend on stage-specific sensitivities. Our study may thereby help to fill a time gap until the results of large-scale randomized clinical trials on the impact of FIT screening on CRC mortality are available, e.g., from the CONFIRM trial (68) (results expected in 2028) and the SCREESCO trial (69) (results expected in 2034).

In summary, this is the first meta-analysis to provide precise estimates of the stage-specific sensitivity of FIT for CRC detection. Although 4 of 5 CRCs of stages II and IV were FIT positive, sensitivity was considerably lower, by approximately 10% points, for stage-I-CRCs. Our results therefore point to the need for further improvement of FITs in detection of early-stage CRC. Such improvement might be achievable by combining FIT with other diagnostic markers. To the best of our knowledge, however, the

previously investigated combinations of FIT with other stool (70) or blood (71) markers achieved—if any—only very limited improvements in overall CRC detection. In conclusion, further large-scale screening cohorts are necessary to obtain more precise stage-specific sensitivity estimates of FIT and to evaluate promising marker combinations to improve the detection of early-stage CRCs.

ACKNOWLEDGEMENTS

We thank Christian Stock for statistical advice.

CONFLICTS OF INTEREST

Guarantor of the article: Tobias Niedermaier, MPH, PhD. The lead author affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Specific author contributions: H.B. designed the study. T.N. and Y.B. conducted the literature search and extracted the data. T.N. conducted the statistical analyses and drafted the manuscript. H.B. and Y.B. contributed important intellectual content and critically revised the manuscript.

Financial support: No specific funding was obtained for this study.

Potential competing interests: None to report.

Study Highlights

WHAT IS KNOWN

- ✓ Overall, FITs detect the majority of CRCs in a single screening round. Cure rates for CRC are highest when detected at an early stage, making it valuable to derive stage-specific estimates of characteristics of FITs for CRC detection.
- ✓ Although FITs are suggested to detect larger proportions of late-stage CRCs compared with early-stage CRCs due to stronger bleeding, these potential differences have not yet been systematically investigated and quantified.

WHAT IS NEW HERE

- ✓ We reviewed and summarized evidence on the stage-specific sensitivity of FITs for CRC detection.
- ✓ Forty-four studies comprising a total of ~92,500 participants and ~3,000 CRC cases were included, yielding stage-specific estimates of FIT accuracy with high precision.
- ✓ Pooled sensitivities (95% CIs) of FIT were 73% (65%–79%) for stage-I-CRCs and 80% (74%–84%), 82% (77%–87%), and 79% (70%–86%) for stages II, III, and IV, respectively. Sensitivity was particularly low for T1 CRCs (40%, 95% CI 21%–64%).
- ✓ These results will enable estimating expected effectiveness and cost-effectiveness of planned screening programs with enhanced accuracy. A comparably low sensitivity of FIT for stage-I-CRCs and for T1 CRCs in particular warrants further research on ways to improve FITs in the detection of early-stage CRC.

REFERENCES

1. Ebell MH, Thai TN, Royalty KJ. Cancer screening recommendations: An international comparison of high income countries. *Public Health Rev* 2018;39:7.
2. Imperiale TF, Gruber RN, Stump TE, et al. Performance characteristics of fecal immunochemical tests for colorectal cancer and advanced adenomatous polyps: A systematic review and meta-analysis. *Ann Intern Med* 2019;170:319–29.
3. American Cancer Society. Survival Rates for Colorectal Cancer, by Stage. (<https://www.cancer.org/cancer/colon-rectal-cancer/detection-diagnosis-staging/survival-rates.html>) (2018). Accessed November 1, 2019.
4. Yoshinaga M, Motomura S, Takeda H, et al. Evaluation of the sensitivity of an immunochemical fecal occult blood test for colorectal neoplasia. *Am J Gastroenterol* 1995;90:1076–9.
5. Sieg A, Scheida M, John MR, et al. Validity of new immunological human fecal hemoglobin and albumin tests in detecting colorectal neoplasms—An endoscopy-controlled study. *Z Gastroenterol* 1998;36:485–90.
6. Wong WM, Lam SK, Cheung KL, et al. Evaluation of an automated immunochemical fecal occult blood test for colorectal neoplasia detection in a Chinese population. *Cancer* 2003;97:2420–4.
7. Young GP, St John DJB, Cole SR, et al. Prescreening evaluation of a brush-based faecal immunochemical test for haemoglobin. *J Med Screen* 2003;10:123–8.
8. Morikawa T, Kato J, Yamaji Y, et al. A comparison of the immunochemical fecal occult blood test and total colonoscopy in the asymptomatic population. *Gastroenterology* 2005;129:422–8.
9. Sohn DK, Jeong SY, Choi HS, et al. Single immunochemical fecal occult blood test for detection of colorectal neoplasia. *Cancer Res Treat* 2005;37:20–3.
10. Li SR, Wang HH, Hu JC, et al. New immunochemical fecal occult blood test with two-consecutive stool sample testing is a cost-effective approach for colon cancer screening: Results of a prospective multicenter study in Chinese patients. *Int J Cancer* 2006;118:3078–83.
11. Smith A, Young GP, Cole SR, et al. Comparison of a brush-sampling fecal immunochemical test for hemoglobin with a sensitive guaiac-based fecal occult blood test in detection of colorectal neoplasia. *Cancer* 2006;107:2152–9.
12. Nakazato M, Yamano H, Matsushita H, et al. Immunologic fecal occult blood test for colorectal cancer screening. *Jpn Med Assoc J* 2006;49:203–7.
13. Hirata I, Hoshimoto M, Saito O, et al. Usefulness of fecal lactoferrin and hemoglobin in diagnosis of colorectal diseases. *World J Gastroenterol* 2007;13:1569–74.
14. Lohsiriwat V, Thavichaigarn P, Awapittaya B. A multicenter prospective study of immunochemical fecal occult blood testing for colorectal cancer detection. *J Med Assoc Thai* 2007;90:2291–5.
15. Mulder SA, van Leerdam ME, van Vuuren AJ, et al. Tumor pyruvate kinase isoenzyme type M2 and immunochemical fecal occult blood test: Performance in screening for colorectal cancer. *Eur J Gastroenterol Hepatol* 2007;19:878–82.
16. Karl J, Wild N, Tacke M, et al. Improved diagnosis of colorectal cancer using a combination of fecal occult blood and novel fecal protein markers. *Clin Gastroenterol Hepatol* 2008;6:1122–8.
17. Shastri YM, Loitsch S, Hoepffner N, et al. Comparison of an established simple office-based immunological FOBT with fecal tumor pyruvate kinase type M2 (M2-PK) for colorectal cancer screening: Prospective multicenter study. *Am J Gastroenterol* 2008;103:1496–504.
18. Graser A, Stieber P, Nagel D, et al. Comparison of CT colonography, colonoscopy, sigmoidoscopy and faecal occult blood tests for the detection of advanced adenoma in an average risk population. *Gut* 2009;58:241–8.
19. Oort FA, Droste J, van der Hulst RWM, et al. Colonoscopy-controlled intra-individual comparisons to screen relevant neoplasia: Faecal immunochemical test vs. guaiac-based faecal occult blood test. *Aliment Pharmacol Ther* 2010;31:432–9.
20. 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.
21. Kalimutho M, Blanco GD, Cretella M, et al. A simplified, non-invasive fecal-based DNA integrity assay and iFOBT for colorectal cancer detection. *Int J Colorectal Dis* 2011;26:583–92.
22. Terhaar sive Droste JS, Oort FA, van der Hulst RW, et al. Higher fecal immunochemical test cutoff levels: Lower positivity rates but still acceptable detection rates for early-stage colorectal cancers. *Cancer Epidemiol Biomarkers Prev* 2011;20:272–80.
23. Tao S, Haug U, Kuhn K, et al. Comparison and combination of blood-based inflammatory markers with faecal occult blood tests for non-invasive colorectal cancer screening. *Br J Cancer* 2012;106:1424–30.
24. Chiu HM, Lee YC, Tu CH, et al. Association between early stage colon neoplasms and false-negative results from the fecal immunochemical test. *Clin Gastroenterol Hepatol* 2013;11:832–8.e1–2.
25. Kaul A, Shah A, Magill FH, et al. Immunological faecal occult blood testing: A discriminatory test to identify colorectal cancer in symptomatic patients. *Int J Surg* 2013;11:329–31.
26. Koga Y, Yamazaki N, Yamamoto Y, et al. Fecal miR-106a is a useful marker for colorectal cancer patients with false-negative results in immunochemical fecal occult blood test. *Cancer Epidemiol Biomarkers Prev* 2013;22:1844–52.
27. Lee YC, Chiu HM, Chiang TH, et al. Accuracy of faecal occult blood test and *Helicobacter pylori* stool antigen test for detection of upper gastrointestinal lesions. *BMJ Open* 2013;3:e003989.
28. Cubiella J, Salve M, Díaz-Ondina M, et al. Diagnostic accuracy of the faecal immunochemical test for colorectal cancer in symptomatic patients: Comparison with NICE and SIGN referral criteria. *Colorectal Dis* 2014;16:O273–82.
29. Imperiale TF, Ransohoff DF, Itzkowitz SH, et al. Multitarget stool DNA testing for colorectal-cancer screening. *N Engl J Med* 2014;370:1287–97.
30. Johnson DA, Barclay RL, Mergener K, et al. Plasma Septin9 versus fecal immunochemical testing for colorectal cancer screening: A prospective multicenter study. *PLoS One* 2014;9:e98238.
31. Kim BC, Joo J, Chang HJ, et al. A predictive model combining fecal calgranulin B and fecal occult blood tests can improve the diagnosis of colorectal cancer. *PLoS One* 2014;9:e106182.

32. Elsafi SH, Alqahtani NI, Zakary NY, et al. The sensitivity, specificity, predictive values, and likelihood ratios of fecal occult blood test for the detection of colorectal cancer in hospital settings. *Clin Exp Gastroenterol* 2015;8:279–84.
33. Baxter NT, Ruffin MT, Rogers MAM, et al. Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. *Genome Med* 2016;8:37.
34. Chen CH, Yan SL, Yang TH et al. The relationship between the methylated septin-9 DNA blood test and stool occult blood test for diagnosing colorectal cancer in Taiwanese people. *J Clin Lab Anal* [Epub ahead of print July 8, 2016.]
35. Symonds EL, Pedersen SK, Baker RT, et al. A blood test for methylated BCAT1 and IKZF1 vs. a fecal immunochemical test for detection of colorectal neoplasia. *Clin Transl Gastroenterol* 2016;7:e137.
36. Wong SH, Kwong TN, Chow TC, et al. Quantitation of faecal Fusobacterium improves faecal immunochemical test in detecting advanced colorectal neoplasia. *Gut* 2016;66:1441–8.
37. Aniwon S, Ratanachu EkT, Pongprasobchai S, et al. The optimal cut-off level of the fecal immunochemical test for colorectal cancer screening in a country with limited colonoscopy resources: A Multi-Center Study from Thailand. *Asian Pac J Cancer Prev* 2017;18:405–12.
38. Brenner H, Niedermaier T, Chen H. Strong subsite-specific variation in detecting advanced adenomas by fecal immunochemical testing for hemoglobin. *Int J Cancer* 2017;140:2015–22.
39. Xie YH, Gao QY, Cai GX, et al. Fecal Clostridium symbiosum for noninvasive detection of early and advanced colorectal cancer: Test and validation studies. *EBioMedicine* 2017;25:32–40.
40. Oono Y, Iriguchi Y, Doi Y, et al. A retrospective study of immunochemical fecal occult blood testing for colorectal cancer detection. *Clin Chim Acta* 2010;411:802–5.
41. Nakama H, Kamijo N, Fujimori K, et al. Relationship between fecal sampling times and sensitivity and specificity of immunochemical fecal occult blood tests for colorectal cancer—A comparative study. *Dis Colon Rectum* 1997;40:781–4.
42. Suehiro Y, Hashimoto S, Higaki S, et al. Blood free-circulating DNA testing by highly sensitive methylation assay to diagnose colorectal neoplasias. *Oncotarget* 2018;9:16974–87.
43. Xie L, Jiang XY, Li Q, et al. Diagnostic value of methylated Septin9 for colorectal cancer detection. *Front Oncol* 2018;8:247.
44. van Turenhout ST, van Rossum LG, Oort FA, et al. Similar fecal immunochemical test results in screening and referral colorectal cancer. *World J Gastroenterol* 2012;18:5397–403.
45. Rex DK, Boland CR, Dominitz JA, et al. Colorectal cancer screening: Recommendations for physicians and patients from the U.S. multi-society task force on colorectal cancer. *Am J Gastroenterol* 2017;112:1016–30.
46. Katsoula A, Paschos P, Haidich AB, et al. Diagnostic accuracy of fecal immunochemical test in patients at increased risk for colorectal cancer: A meta-analysis. *JAMA Intern Med* 2017;177:1110–8.
47. Robertson DJ, Lee JK, Boland CR, et al. Recommendations on fecal immunochemical testing to screen for colorectal neoplasia: A consensus statement by the US multi-society task force on colorectal cancer. *Gastroenterology* 2017;152:1217–1237.e3.
48. 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;45:703–12.
49. Hernandez V, Cubiella J, Gonzalez-Mao MC, et al. Fecal immunochemical test accuracy in average-risk colorectal cancer screening. *World J Gastroenterol* 2014;20:1038–47.
50. 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;160:171–81.
51. 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;155:529–36.
52. 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.
53. Macaskill P, Walter SD, Irwig L. A comparison of methods to detect publication bias in meta-analysis. *Stat Med* 2001;20:641–54.
54. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria, 2016.
55. Doebler P, Holling H. Meta-analysis of diagnostic accuracy with mada. (<http://cran.r-project.org/web/packages/mada/>) 2017. Accessed November 1, 2019.
56. Reitsma JB, Glas AS, Rutjes AW, et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;58:982–90.
57. Wakamura K, Kudo SE, Miyachi H, et al. Characteristics of colorectal tumours in asymptomatic patients with negative immunochemical faecal occult blood test results. *Gastroenterology* 2015;148:S744.
58. Chang LC, Shun CT, Hsu WF, et al. Fecal immunochemical test detects sessile serrated adenomas and polyps with a low level of sensitivity. *Clin Gastroenterol Hepatol* 2017;15:872–9.e1.
59. Niedermaier T, Weigl K, Hoffmeister M, et al. Diagnostic performance of flexible sigmoidoscopy combined with fecal immunochemical test in colorectal cancer screening: meta-analysis and modeling. *Eur J Epidemiol* 2017;32:481–93.
60. Väyrynen JP, Tuomisto A, Väyrynen SA, et al. Preoperative anemia in colorectal cancer: Relationships with tumor characteristics, systemic inflammation, and survival. *Sci Rep* 2018;8:1126.
61. Crooke H, Kobayashi M, Mitchell B, et al. Estimating 1- and 5-year relative survival trends in colorectal cancer (CRC) in the United States: 2004 to 2014. *J Clin Oncol* 2018;36:587.
62. Amitay EL, Cuk K, Niedermaier T, et al. Factors associated with false-positive fecal immunochemical tests in a large German colorectal cancer screening study. *Int J Cancer* 2018;144:2419–2427.
63. Gies A, Cuk K, Schrotz-King P, et al. Direct comparison of diagnostic performance of 9 quantitative fecal immunochemical tests for colorectal cancer screening. *Gastroenterology* 2018;154:93–104.
64. Leeflang MM, Bossuyt PM, Irwig L. Diagnostic test accuracy may vary with prevalence: Implications for evidence-based diagnosis. *J Clin Epidemiol* 2009;62:5–12.
65. Willis BH. Spectrum bias—Why clinicians need to be cautious when applying diagnostic test studies. *Fam Pract* 2008;25:390–6.
66. 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.
67. Cole SR, Tucker GR, Osborne JM, et al. Shift to earlier stage at diagnosis as a consequence of the National Bowel Cancer Screening Program. *Med J Aust* 2013;198:327–30.
68. Dominitz JA, Robertson DJ, Ahnen DJ, et al. Colonoscopy vs. fecal immunochemical test in reducing mortality from colorectal cancer (CONFIRM): Rationale for study design. *Am J Gastroenterol* 2017;112:1736–46.
69. Ribbing Wilén H, Blom J, Höjjer J, et al. Fecal immunochemical test in cancer screening—Colonoscopy outcome in FIT positives and negatives. *Scand J Gastroenterol* 2019;54:303–10.
70. Niedermaier T, Weigl K, Hoffmeister M, et al. Fecal immunochemical tests combined with other stool tests for colorectal cancer and advanced adenoma detection: A systematic review. *Clin Transl Gastroenterol* 2016;7:e175.
71. Niedermaier T, Weigl K, Hoffmeister M, et al. Fecal immunochemical tests in combination with blood tests for colorectal cancer and advanced adenoma detection-systematic review. *United European Gastroenterol J* 2018;6:13–21.

This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.