Published in final edited form as:
Ann Intern Med. 2014 February 4; 160(3): 171. doi:10.7326/M13-1484.

Accuracy of Fecal Immunochemical Tests for Colorectal Cancer: Systematic Review and Meta-analysis

Jeffrey K. Lee, MD, MAS, **Elizabeth G. Liles, MD, MCR**, **Stephen Bent, MD**, **Theodore R. Levin, MD**, and **Douglas A. Corley, MD, PhD**
University of California, San Francisco, San Francisco, California; Center for Health Research, Kaiser Permanente Northwest, Portland, Oregon; and Kaiser Permanente Northern California, Oakland, California

Abstract

Background—Performance characteristics of fecal immunochemical tests (FITs) to screen for colorectal cancer (CRC) have been inconsistent.

Purpose—To synthesize data about the diagnostic accuracy of FITs for CRC and identify factors affecting its performance characteristics.

Data Sources—Online databases, including MEDLINE and EMBASE, and bibliographies of included studies from 1996 to 2013.

Study Selection—All studies evaluating the diagnostic accuracy of FITs for CRC in asymptomatic, average-risk adults.

Data Extraction—Two reviewers independently extracted data and critiqued study quality.

Requests for Single Reprints: Douglas A. Corley, MD, PhD, Kaiser Permanente Division of Research, 2000 Broadway, Oakland, CA 94612; Douglas.Corley@kp.org.
Current Author Addresses: Dr. Lee: University of California, San Francisco, 513 Parnassus Avenue, S-357, San Francisco, CA 94143.
Dr. Liles: Center for Health Research, Kaiser Permanente, 3800 North Interstate Avenue, Portland, OR 97227-1110.
Dr. Bent: Osher Center for Integrative Medicine, Department of Medicine, University of California, San Francisco Veterans Affairs Medical Center 111-A1, 4150 Clement Street, San Francisco, CA 94121.
Drs. Levin and Corley: Kaiser Permanente Division of Research, 2000 Broadway, Oakland, CA 94612.

Potential Conflicts of Interest: Disclosures can be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M13-1484.

Current author addresses and author contributions are available at www.annals.org.

Author Contributions: Conception and design: J.K. Lee, E.G. Liles, T.R. Levin, D.A. Corley.
Analysis and interpretation of the data: J.K. Lee, E.G. Liles, S. Bent.
Drafting of the article: J.K. Lee, E.G. Liles.
Critical revision of the article for important intellectual content: J.K. Lee, E.G. Liles, S. Bent, T.R. Levin, D.A. Corley.
Final approval of the article: J.K. Lee, E.G. Liles, S. Bent, T.R. Levin, D.A. Corley.
Provision of study materials or patients: J.K. Lee.
Statistical expertise: J.K. Lee, S. Bent.
Obtaining of funding: J.K. Lee.
Administrative, technical, or logistic support: J.K. Lee, E.G. Liles.
Collection and assembly of data: J.K. Lee, E.G. Liles.

See also:
Web-Only
Supplement

Data Synthesis—Nineteen eligible studies were included and meta-analyzed. The pooled sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio of FITs for CRC were 0.79 (95% CI, 0.69 to 0.86), 0.94 (CI, 0.92 to 0.95), 13.10 (CI, 10.49 to 16.35), 0.23 (CI, 0.15 to 0.33), respectively, with an overall diagnostic accuracy of 95% (CI, 93% to 97%). There was substantial heterogeneity between studies in both the pooled sensitivity and specificity estimates. Stratifying by cutoff value for a positive test result or removal of discontinued FIT brands resulted in homogeneous sensitivity estimates. Sensitivity for CRC improved with lower assay cutoff values for a positive test result (for example, 0.89 [CI, 0.80 to 0.95] at a cutoff value less than 20 μg/g vs. 0.70 [CI, 0.55 to 0.81] at cutoff values of 20 to 50 μg/g) but with a corresponding decrease in specificity. A single-sample FIT had similar sensitivity and specificity as several samples, independent of FIT brand.

Limitations—Only English-language articles were included. Lack of data prevented complete subgroup analyses by FIT brand.

Conclusion—Fecal immunochemical tests are moderately sensitive, are highly specific, and have high overall diagnostic accuracy for detecting CRC. Diagnostic performance of FITs depends on the cutoff value for a positive test result.

Primary Funding Source—National Institute of Diabetes and Digestive and Kidney Diseases and National Cancer Institute.

Colorectal cancer (CRC) is the second-leading cause of cancer-related deaths in the United States (1). Randomized, controlled trials have shown that annual or biennial fecal occult blood tests (FOBTs) are associated with a 15% to 33% decrease in CRC mortality rates (2–4). However, FOBTs only detect approximately 13% to 50% of cancer with 1 round of screening in asymptomatic patients (5, 6). In addition, adherence to repeated rounds of FOBTs in real-world screening programs is low, raising concern about their effectiveness as screening tests (7, 8).

Fecal immunochemical tests (FITs) are more sensitive at detecting both CRC and adenomas than FOBTs (9, 10). Many FITs require only 1 or 2 stool samples, and none require dietary or medication restrictions, increasing ease of use. In 2008, several U.S. professional societies endorsed the use of FITs to replace FOBTs because of the former's improved performance characteristics and potential for higher participation rates (10, 11). Countries in Europe and Asia have also adopted widespread CRC screening programs using FITs (12, 13). However, the diagnostic characteristics of these tests have been difficult to estimate, with reported sensitivities ranging from 25% to 100% for CRC and specificities usually exceeding 90% (9, 14, 15). The lack of a precise estimate of sensitivity has resulted in confusion among health care providers about the sources of this variation, how best to apply FITs for CRC screening, the optimal number of stool samples for testing, optimal cutoff value for a positive test result, and whether any brand of FIT is superior to others. Our analysis provides a quantitative meta-analysis of the diagnostic accuracy (sensitivity and specificity) of FITs for CRC. In addition, we explored potential sources of heterogeneity by analyzing subgroups classified by FIT sample number, cutoff value for a positive test result, FIT brand, and the reference standard.

Methods

We developed a protocol on the basis of standard guidelines for the systematic review of diagnostic studies (16, 17) and the strategy used for the U.S. Preventive Services Task Force review in 2008 (9). We followed the STARD (Standards for the Reporting of Diagnostic Accuracy Studies) (18) and Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (19) statements for reporting our systematic review. This study was conducted as part of the National Cancer Institute–funded consortium, Population-Based Research Optimizing Screening through Personalized Regimens. The overall aim of this consortium is to conduct multisite, coordinated, transdisciplinary research to evaluate and improve cancer screening processes.

Data Sources and Searches

We included all studies identified in the previous USPSTF report (9) plus other studies identified by a search of FIT for CRC between 1 January 2008 and 31 August 2013 using MEDLINE (via Ovid), EMBASE, Database of Abstracts of Reviews of Effects, Health Technology Assessment Database, Cochrane Database of Systematic Reviews, and Cochrane Central Register of Controlled Trials. We also searched bibliographies and reference lists of eligible papers and related reviews, consulted experts in the field, and contacted several authors from the included studies to locate additional studies. The Appendix Table 1 (available at www.annals.org) provides further details of our search strategy.

Study Selection

Two persons independently reviewed the pertinent studies to determine eligibility. We included studies if they met all of the following criteria: evaluated the diagnostic accuracy of FITs for CRC; reported absolute numbers of true-positive, false-negative, true-negative, and false-positive observations, or if these same variables could be obtained from personal communication; used a randomized trial or cohort study design; evaluated adult participants who were asymptomatic and older than 18 years with a mean age greater than 40 years; and reported an appropriate reference standard (colonoscopy or ≥2-year longitudinal follow-up of controls with medical records or cancer registry). Given that only a subset of studies reported data on adenomatous polyps and that there is variability in definitions of polyps, we limited the scope of this analysis to test performance characteristics for detecting CRC; we excluded studies reporting test performance estimates for detection of adenomas only. We did not include conference abstracts and case– control studies, which, by creating spectrum bias, can overestimate the accuracy of a test (20). To avoid duplicate reporting of the same population for studies reporting several cutoff values or numbers of samples, we used the cutoff value or sample number most commonly used in current practice in the United States, used in national recommendations, or recommended by expert opinion in the main analyses. In addition, we selected the sample number or cutoff value a priori that was most similar to those in other studies for our subgroup analyses.

Data Extraction and Quality Assessment

Two reviewers independently evaluated and extracted relevant information from each included study and assessed study quality via the Quality Assessment of Diagnostic Accuracy Studies 2 instrument (21). For studies with incomplete or unavailable information, we contacted the corresponding authors or coauthors to complete missing information. Of the 15 contacted authors, 12 provided additional data. We converted units for cutoff thresholds for a positive test result in each study to micrograms of hemoglobin per gram of stool, as recommended by leading experts (22).

Data Synthesis and Analysis

We calculated the sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio (LR), and negative LR with 95% CI of each study. A positive LR greater than 5 and a negative LR less than 0.2 provide strong diagnostic evidence to rule in or rule out diagnoses, respectively (23).

The overall pooled sensitivity and specificity of FIT for CRC were estimated using a bivariate random-effects model (24). We calculated the pooled positive LR and negative LR along with the respective CI using the bivariate model (24) according to the method used by Zwinderman and colleagues (25). We also generated a hierarchical summary receiver-operating characteristic curve that plots the individual and summary estimates of sensitivities and specificities along with a 95% confidence and prediction region (26). Last, we calculated the area under the hierarchical summary receiver-operating characteristic curve. An area under the curve between 0.9 and 1.0 indicates that the test in question is highly accurate (27).

The Q value and the inconsistency index (I^2) test were used to estimate the heterogeneity between each study (28). We regarded values of 25%, 50%, and 75% for the I^2 as indicative of low, moderate, and high statistical heterogeneity, respectively (28). In addition, we calculated the between-study variance of logit sensitivity and logit specificity (24, 29). In diagnostic accuracy studies, 1 of the primary causes of heterogeneity is the threshold effect, which occurs when different cutoff values are used between studies to define a positive (or negative) test result. We searched for evidence of a threshold effect by calculating the squared correlation coefficient estimated from the between-study covariance variable in the bivariate model (30).

We stratified studies into 4 subgroups on the basis of the number of FIT samples (1, 2, or 3 samples), prespecified cutoff values of fecal hemoglobin concentration for a positive test result (<20 μg/g, 20 to 50 μg/g, and >50 μg/g), brand, and reference standard used to follow up on patients with negative FIT results. Cutoff values were grouped to ensure an adequate number of data sets for each analysis. To determine whether studies using older (discontinued) FITs were causing heterogeneity in our summary estimates, we did sensitivity analyses by removing these studies and recalculating the I^2 test for the remaining group. In addition to threshold effect and subgroup analyses, we did a bivariate random-effects meta-regression analysis to identify additional sources of heterogeneity that may have influenced our overall summary estimates (30). We used the following prespecified

variables for our meta-regression: type of FIT (qualitative, point-of-care tests or quantitative, automated tests), geographic region (Asian or non-Asian countries), and enrollment of patients younger than 40 years. We used Stata, version 12.0 (StataCorp, College Station, Texas), for all statistical analyses. All tests were 2-sided, and we considered *P* values less than 0.05 to be statistically significant.

Role of the Funding Source

The study was funded by the National Institute of Diabetes and Digestive and Kidney Diseases and the National Cancer Institute. The funding source had no role in the conception, design, analysis, or conduct of the review.

Results

Study Selection

The 2008 USPSTF report (9) included 9 studies in its systematic review (31–39); our literature search identified 1771 additional new potential sources (Figure 1). After abstract review, we identified 53 articles for full-text review; of these, 18 unique articles satisfied all inclusion criteria and were included in our analysis (14, 15, 31–46). Because 1 article (46) evaluated more than 1 FIT brand in a head-to-head comparison, the final analysis included 19 studies or data sets.

Characteristics of Included Studies

Table 1 and the Supplement (available at www.annals.org) show the main characteristics of the included studies. Eighteen articles described 19 cohort studies of FIT sensitivity and specificity for CRC in average-risk asymptomatic patients; sample sizes ranged from 80 to 27 860. Twelve studies (14, 33–36, 40–42, 44–46) used colonoscopy in all patients, regardless of FIT results, as the reference standard, allowing for detection of both adenomatous polyps and CRC (although the current analysis used only each study's cancer end point because of heterogeneity of the polyp data). Seven studies (15, 31, 32, 37–39, 43) used colonoscopy in patients with positive FIT results, with longitudinal follow-up of patients with negative FIT results through cancer registries, medical records, national insurance claims, or telephone contact approximately 2 years later as the reference standard. Ten studies (14, 34–37, 39–41, 44, 45) were done in Asian countries, and 9 studies (15, 31–33, 38, 42, 43, 46) were done in non-Asian countries. The mean age of the study populations ranged from 45.2 to 62.7 years, with 5 studies (14, 34, 35, 39, 45) having a mean age between 40 and 50 years. Although the settings for recruitment of patients varied from a general hospital to community-based clinics, all studies recruited asymptomatic patients without a history of CRC or previous screening and often excluded those with a known history of a medical condition (that is, a genetic syndrome or inflammatory bowel disease) that would increase risk for CRC. Funding sources for individual studies varied: Five studies reported nonindustry funding only (36, 37, 42, 44, 45), 2 studies reported partial industry funding (15, 33), 6 studies reported nonindustry funding except for provision of the FIT kits by the manufacturer (31, 32, 38, 41, 43, 46), and 5 studies did not report the funding source (14, 34, 35, 39, 40).

The meta-analysis included 8 different FITs (OC-Micro/Sensor, which involves the combination of OC-Micro [Polymedco, Cortlandt Manor, New York] and OC-Sensor [Eiken Chemical, Tokyo, Japan]; OC-Light [Eiken Chemical]; OC-Hemodia [Eiken Chemical]; Monohaem [Nihon Pharmaceutical, Japan]; MagStream HemSp [Fujirebio, Tokyo, Japan]; FlexSure OBT/Hemoccult ICT [Beckman Coulter, Fullerton, California]; HemeSelect/Immudia HemSp [Fujirebio]; and Ridascreen Haemoglobin [R-Biopharm AG, Darmstadt, Germany]). However, 2 FITs (OC-Hemodia and HemeSelect/Immudia HemSp) have been discontinued and are no longer produced in the United States. Eight studies evaluated 1 of the qualitative FITs (OC-Light [34, 43–45], Monohaem [36, 37], FlexSure OBT/Hemoccult ICT [32], and Hemeselect/Immudia HemSp [31]) for which the cutoff fecal hemoglobin concentration for a positive test result was preset and the test was read as positive or negative. Eleven studies evaluated 1 of the quantitative FITs (OC-Micro/Sensor [15, 33, 41, 42, 46], OC-Hemodia [14, 39, 40], MagStream HemSp [35, 38], and Ridascreen Haemoglobin [46]) for which the cutoff fecal hemoglobin concentration for a positive test result could be adjusted by the end user. Six studies (15, 31–33, 38, 40) analyzed the performance characteristics of several FIT samples, whereas the remaining 13 studies (14, 34–37, 39, 41–46) used only a single-sample FIT assay. All studies evaluating several FIT samples defined 1 or more positive samples (meeting the cutoff threshold) as a positive result of the test overall. The cutoff value for a positive FIT result varied widely, ranging from 6.1 to 300.0 μg/g; however, 14 studies (74%) reported a cutoff between 10 and 20 μg/g (14, 15, 33, 34, 36, 37, 39–46).

Quality Assessment

The Quality Assessment of Diagnostic Accuracy Studies 2 instrument (21) (Appendix Figure 1 and Appendix Table 2, available at www.annals.org) suggested the greatest risk of bias occurred in the “flow and timing” and “reference standard” categories. This is mainly because of the 7 studies that used variable reference standards depending on the FIT results and because endoscopists were not blinded to the FIT results (15, 31, 32, 37–39, 43). The greatest concern of applicability came from the “patient selection” category, where 7 studies included some patients who were either younger than 40 years or older than 80 years (14, 32, 34, 35, 37, 40, 45); 2 studies included some patients with a family history of CRC (33, 42), and 1 study had no information to determine mean age (33).

Overall Accuracy of FIT

The overall pooled sensitivity, specificity, positive LR, and negative LR of FITs for CRC were 0.79 (95% CI, 0.69 to 0.86), 0.94 (CI, 0.92 to 0.95), 13.10 (CI, 10.49 to 16.35), and 0.23 (CI, 0.15 to 0.33), respectively (Figure 2 and Appendix Figure 2, available at www.annals.org). The overall accuracy of FIT was 95% (CI, 93% to 97%) (Appendix Figure 3, available at www.annals.org).

Investigation of Heterogeneity

We found substantial heterogeneity between studies when calculating the pooled sensitivity (I^2 = 68.45%), specificity (I^2 = 98.50%), positive LR (I^2 = 89.76%), and negative LR (I^2 = 77.78%) of FITs for CRC using the bivariate model (Figure 2 and Appendix Figure 2).

Excluding studies that used discontinued FITs reduced heterogeneity for sensitivity (I^2 = 51.0%), specificity (I^2 = 98.1%), positive LR (I^2 = 88.0%), and negative LR (I^2 = 42.4%). Excluding discontinued FITs also moderately increased the pooled sensitivity (0.82 [CI, 0.73 to 0.89]) and decreased the negative LR (0.19 [CI, 0.12 to 0.30]) estimates without changing the pooled specificity estimates (0.94 [CI, 0.92 to 0.95]). The percentage of heterogeneity likely because of a threshold effect was 43%, suggesting that the use of different cutoffs for a positive test result between studies contributed substantially to the heterogeneity of our overall pooled estimates.

Subgroup Analysis

Number of FIT Samples—The pooled performance characteristics of FIT for CRC were similar regardless of the number of stool samples tested with FITs (Table 2 and Appendix Figures 4 to 6). The pooled sensitivities for 1-, 2-, and 3-sample FITs were 0.78 (CI, 0.65 to 0.87), 0.77 (CI, 0.59 to 0.89), and 0.80 (CI, 0.66 to 0.89), respectively (Table 2 and Appendix Figures 4 to 6, available at www.annals.org). The pooled specificities for 1-, 2-, and 3-sample FITs were 0.95 (CI, 0.93 to 0.96), 0.93 (CI, 0.90 to 0.95), and 0.93 (CI, 0.89 to 0.95), respectively (Table 2). The pooled negative LRs for 1-, 2-, and 3-sample FITs were 0.24 (CI, 0.15 to 0.38), 0.25 (CI, 0.13 to 0.49), and 0.21 (CI, 0.12 to 0.38), respectively (Table 2). Aside from the 3-sample FIT, we saw high heterogeneity in the pooled 1- and 2-sample FIT sensitivities and negative LRs (Table 2). We also saw significant heterogeneity in the pooled 1-, 2-, and 3-sample FIT specificities and positive LRs. When we removed discontinued FITs in our 1-sample FIT subgroup, pooled sensitivity (0.78 [CI, 0.65 to 0.87]) and negative LR (0.23 [CI, 0.14 to 0.38]) estimates remained similar, but heterogeneity decreased from an I^2 of 75.9% to 58.1% for sensitivity and 84.8% to 49.8% for negative LR (Tables 2 and 3). However, specificity and the positive LR estimates and their associated heterogeneity remained unchanged. We could not do a sensitivity analysis in our 2-sample FIT subgroup because of the lack of data sets or studies.

Cutoff Value for a Positive FIT Test—Varying cutoff values used to define an abnormal test influenced the performance characteristics of FIT for CRC. Sensitivity decreased with increasing cutoff values, from 0.86 (CI, 0.75 to 0.92) at cutoff values less than 20 μg/g to 0.67 (CI, 0.59 to 0.74) at cutoff values greater than 50 μg/g (Table 2 and Appendix Figures 7 to 9, available at www.annals.org). However, specificity increased from 0.91 (CI, 0.89 to 0.93) at cutoff values less than 20 μg/g to 0.96 (CI, 0.94 to 0.98) at those greater than 50 μg/g. The negative LR decreased with decreasing cutoff values, with those less than 20 μg/g (negative LR, 0.16 [CI, 0.09 to 0.28]) showing the strongest diagnostic evidence to rule out CRC among the 3 cutoff groups (Table 2). The positive LRs of FITs at all 3 cutoff subgroups were sufficiently high to be qualified as a rule-in diagnostic tool, whereas the negative LR of FIT at cutoff values 20 to 50 μg/g and greater than 50 μg/g were not low enough to be used as a rule-out screening test for CRC (Table 2). We saw high heterogeneity at cutoff values less than 20 μg/g and 20 to 50 μg/g for sensitivity and negative LR and in all cutoffs for specificity and positive LR (Table 2 and Appendix Figures 7 to 9). When we removed discontinued FITs, pooled sensitivity estimates improved from 0.86 to 0.89 (CI, 0.80 to 0.95) at cutoff values less than 20 μg/g and were more homogeneous (I^2 = 26.4%) (Table 3). Similarly, pooled sensitivity estimates improved from

0.63 to 0.70 (CI, 0.55 to 0.81) at cutoff values 20 to 50 μg/g and were also more homogeneous (I^2 = 0%) after removing discontinued FITs (Table 3). More homogeneous negative LR estimates were also seen at cutoff values less than 20 μg/g and 20 to 50 μg/g after removal of discontinued FITs (Table 3). However, specificity and positive LR estimates remained similar at cutoff values less than 20 μg/g and 20 to 50 μg/g despite removal of discontinued FITs with high heterogeneity (Table 3).

FIT Brand—There were no considerable differences in performance characteristics among various commercial FIT brands, although the CIs were fairly wide for sensitivity (Figure 2 and Table 1). Only 2 FIT brands (OC-Micro/Sensor and OC-Light) had several studies that could be pooled in our subgroup analyses. The pooled sensitivity of OC-Light (0.93 [CI, 0.83 to 0.97]) was fairly similar to OC-Micro/Sensor (0.86 [CI, 0.68 to 0.95]) (Table 2 and Appendix Figures 10 and 11, available at www.annals.org). Similarly, there was no difference in specificity between OC-Light (0.91 [CI, 0.88 to 0.92]) and OC-Micro/Sensor (0.91 [CI, 0.87 to 0.94]). In both tests, heterogeneity between studies was low for sensitivity (OC-Light, I^2 = 26.6%; OC-Micro, I^2 = 0%) but high for specificity (Table 2 and Appendix Figures 10 and 11). The negative LR estimates were less than 0.20 for both tests (OC-Micro/Sensor and OC-Light) with low heterogeneity, indicating the strong diagnostic ability to rule out CRC (Table 2). Four studies of OC-Light included a total of 20 785 participants, with study sample sizes ranging from 1756 to 8822 (34, 43–45); 3 of the 4 studies occurred in Taiwan (34, *44*, 45), using colonoscopy in all participants as the reference standard. All 4 studies (34, 43–45) used a single sample with the same cutoff value for a positive test (10 μg/g). In the 5 studies of OC-Micro/Sensor, a total of 5545 participants were enrolled, with study sample sizes ranging from 80 to 2235 (15, 33, 41, 42, 46); 2 studies were done in Israel (15, 33), 1 in Korea (41), 1 in the Netherlands (42), and 1 in Germany. Four studies used colonoscopy as the reference standard (33, 41, 42, 46) as compared with Levi and colleagues' study (15), which used a 2-year longitudinal follow-up for FIT-negative patients. Three of the 5 studies used a 3-sample FIT with cutoffs ranging from 14 to 15 μg/g (15, 33, 41). The fourth and fifth study used a single-sample FIT with cutoffs ranging from 6.1 to 15.0 μg/g (42, 46).

Reference Standard—The sensitivity estimates varied depending on the reference standard used to follow up on patients with negative FIT results. Studies using a colonoscopy to follow up on patients with negative FIT results had a lower sensitivity (0.71 [CI, 0.58 to 0.81]) compared with studies using 2-year or longer longitudinal follow-up (0.87 [CI, 0.75 to 0.93]) (Table 2 and Appendix Figures 12 and 13, available at www.annals.org). Specificity remained similar for both subgroups. We saw low heterogeneity between studies that used a 2-year longitudinal follow-up for sensitivity and negative LR estimates (Table 2). In contrast, we saw high heterogeneity among studies that used colonoscopy to follow up on patients with negative FIT results for both sensitivity and negative LR. However, after removing discontinued FITs, sensitivity estimates for the colonoscopy studies improved from 0.73 to 0.77 (CI, 0.65 to 0.86) and were more homogeneous (I^2 = 45.7%) (Table 3).

Meta-regression—We did a meta-regression analysis to evaluate for additional causes of heterogeneity. The meta-regression showed that geographic region (Asian countries or non-Asian countries) and the type of FIT (qualitative or quantitative) were also significant predictors of heterogeneity for sensitivity (Appendix Table 3, available at www.annals.org). In addition, all prespecified covariates were significant predictors of heterogeneity for specificity; however, the magnitude of change between the summary estimates and CIs in each subgroup was small.

Discussion

Our meta-analysis suggests that the pooled sensitivity and specificity of FITs for CRC were approximately 79% and 94%, respectively. In addition, the overall accuracy of FIT was 95%. We saw high heterogeneity among studies (as defined by the I^2 statistic) for all diagnostic measures. To address this issue, we did prespecified subgroup analyses to investigate potential sources of heterogeneity between studies.

We assumed that the heterogeneity in test performance may be because of the number of FIT samples used in each study. Surprisingly, we found that increasing the number of FIT samples did not affect the pooled sensitivities, specificities, positive LRs, or negative LRs of FITs for CRC. Only 2 studies (36, 41) to date have directly evaluated the effect of FIT sample number on the diagnostic accuracy of FITs in average-risk asymptomatic participants; these showed a tradeoff between sensitivity and specificity with increasing FIT samples. However, in both studies, the magnitude of change between 1-, 2-, and 3-sample FIT accuracy was highly variable depending on the cutoff value and reference standard used. A 1-sample FIT regimen for CRC detection may ultimately be desirable, given the importance of optimizing overall adherence in repeated rounds of annual or biennial testing for programmatic screening.

We also evaluated whether different cutoff values for a positive FIT result led to the disparate results found between previous studies; these analyses confirmed that the sensitivity and specificity of FIT for CRC were influenced by the varying cutoff values used to define an abnormal test. Not surprisingly, sensitivity improved with lower assay cutoff values for a positive test result but with a corresponding decrease in specificity. More important, sensitivity estimates were more homogeneous after being stratified by cutoff and discontinued FITs (that is, OC-Hemodia) were removed suggesting that 89%, 70%, and 67% are reasonable estimations of the true sensitivity at cutoff values less than 20 μg/g and 20 to 50 μg/g and greater than 50 μg/g, respectively. Although our study could not identify the optimal cutoff value for CRC screening, a FIT cutoff value less than 20 μg/g had the best combination of sensitivity and specificity for CRC (89% and 91%, respectively) and the lowest negative LR (0.16) compared with the subgroups with cutoff values of 20 to 50 μg/g and greater than 50 μg/g. However, resource use is an important consideration when choosing a cutoff threshold because studies using a 1-sample FIT with cutoff values less than 20 μg/g had positivity rates from 5.3% to 14.2%, which were generally greater than those for 1-sample FITs at cutoff values of 20 to 50 μg/g (1.4% to 7.5%). Considering the lack of colonoscopy resources across the world, identifying an optimal cutoff value for defining a positive result deserves considerable attention because this number can influence

both the number of cancer cases detected as well as the number of colonoscopies needed in a CRC screening program.

Overall, no single commercial FIT brand seemed to perform markedly better or worse than others for CRC detection, but this finding should be interpreted cautiously because most studies did not include head-to-head comparisons. Only 2 of the 8 FIT brands had several studies for pooling in our subgroup analysis by brand, which demonstrated that OC-Light had similar sensitivity and specificity (93% and 91%, respectively) compared with OC-Micro/Sensor (86% and 91%). Furthermore, there was low statistical heterogeneity in both FIT brands for sensitivity. Although Eiken manufactures both tests, OC-Light is a qualitative FIT that is primarily used as a rapid point-of-care test. In contrast, OC-Micro/Sensor is a quantitative FIT that requires a machine or analyzer to measure human hemoglobin through an optical latex agglutination technique. Of interest, studies using quantitative FITs had a lower sensitivity compared with qualitative FITs (73% vs. 85%) in our meta-regression. However, when removing the discontinued quantitative FIT, OC-Hemodia, sensitivity in the quantitative FIT group increased from 73% to 77%, with more homogeneous results (I^2 = 12.1%), indicating that the lower sensitivity of OC-Hemodia was probably the main factor contributing to the variation between quantitative and qualitative FIT sensitivity estimates. The relatively similar sensitivity and specificity between the 2 FIT brands (OC-Light and OC-Micro/Sensor) and FIT formats is clinically useful for CRC screening programs interested in fecal-based screening options. Given the convenience, costs, and rapid turnaround, qualitative FITs, such as OC-Light, may be better suited for use in clinics, whereas quantitative FITs may be more efficient and cost-effective when used in large, organized screening programs.

Our study had several limitations. First, heterogeneity existed in most analyses. Nonetheless, the more homogeneous subgroup summary estimates were generally very similar to the overall summary estimates, suggesting that despite some statistical heterogeneity, which was probably from the high degree of precision of each study, the overall summary measures are reasonable estimations of overall FIT test accuracy. Second, our meta-analysis was subject to the detection, verification, and spectrum biases of the original studies. Of the 19 included studies, 7 used at least a 2-year longitudinal follow-up for patients with negative FIT results as an acceptable reference standard. Previous studies have indicated that using a 2- or 3-year follow-up as a reference standard potentially overestimates sensitivity and underestimates specificity because of verification bias (9, 36). Nakama and colleagues (36) found that sensitivity of FITs for CRC using a 2-year longitudinal follow-up was 83% compared with 71% at 3 years. Whether this difference in sensitivity is because of incident, rapidly growing cancer between 2 and 3 years of follow-up or verification bias is unclear. However, cancer is even missed when colonoscopy is used at a rate of at least 0.2% to 5.0% (47–53). Third, 4 studies enrolling patients younger than 40 years may have introduced spectrum bias into our analysis. However, this factor only contributed to the heterogeneity in specificity. Fourth, we omitted non-English studies, which could have resulted in language bias, although previous studies suggest that this has little effect on the overall conclusions (54, 55). Fifth, we could not determine the sensitivity and specificity of FIT for CRC stratified by site (proximal vs. distal) because many studies did not report the site of each CRC. Sixth, our

study may be subject to publication bias. Last, because of the complexity of accounting and adjusting for various definitions of advanced adenoma, the current study does not report on the performance of FIT for advanced adenoma.

In summary, this systematic review and meta-analysis suggests that FITs have high accuracy, high specificity, and moderately high sensitivity for detection of CRC. Pooled test performance estimates show that the accuracy of a specific qualitative FIT (OC-Light) is similar to that of a specific quantitative FIT (OC-Micro/Sensor) for detecting CRC. This finding suggests that FIT type could be customized to different-sized care settings without significant variability in accuracy. In addition, FIT's diagnostic performance was dependent on the cutoff value used to define a positive test. Health systems wishing to optimize use of a quantitative FIT should consider the tradeoff between increasing sensitivity (by lowering the cutoff threshold for a positive test) and the resulting increase in the number of positive test results, which will have a greater effect on colonoscopy resources. The current data do not provide definitive information about the effect of sample number on FIT performance; systems should look at individual studies comparing test performance and positivity rates among 1-, 2-, and 3-sample iterations of the same test (36, 41) to make decisions about performance versus test positivity tradeoffs (for specific FITs) in varying sample number.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank James Allison, MD, for his expert review of our manuscript and Leslie Bienen, DVM, MFA, for her editing of the manuscript.

Grant Support: By the National Institutes of Health (T32DK007007), National Cancer Institute (U54 CA163262), National Institute of Diabetes and Digestive and Kidney Diseases (T32DK007007), National Cancer Institute Cancer Research Network (U24 CA171524), and Population-Based Research Optimizing Screening through Personalized Regimens (U54 CA163262).

Appendix

Appendix Table 1
Search Strategy

MEDLINE (Ovid) Search Strategy
1 fecal.ti,ab,hw. (38014)
2 faecal.ti,ab,hw. (16604)
3 feces.ti,ab,hw. (73832)
4 1 or 2 or 3 (102056)
5 exp immunoassay/ (403733)
6 exp enzyme-linked immunosorbent assay/ (115110)
7 immunochemi$.ti,ab,hw. (26200)
8 exp Immunochemistry/ (238397)
9 or/5-8 (617537)

10 4 and 9 (4973)

11 FIT.ti,ab,hw. (66545)

12 guaiac.ti,ab,hw. (637)

13 occult blood.ti,ab,hw. (5627)

14 fobt$.ti,ab. (840)

15 fob$.ti,ab. (1502)

16 ifobt.ti,ab. (55)

17 ifob$.ti,ab,hw. (58)

18 or/7,10-17 (103072)

19 Insure.mp. (2487)

20 Inform.mp. (25162)

21 19 or 20 (27634)

22 18 and 21 (317)

23 Instant-view.mp. (5)

24 Instant View.mp. (5)

25 Hemoccult.mp. (520)

26 Immocare.mp. (3)

27 Flexsure.mp. (28)

28 Monohaem.mp. (33)

29 Hemosure.mp. (1)

30 Occultech.mp. (1)

31 Quickvue.mp. (89)

32 Clearview.mp. (86)

33 Hemoquant.mp. (47)

34 Hema screen.mp. (2)

35 Hema-screen.mp. (2)

36 Innovacon.mp. (1)

37 OC-Micro.mp. (9)

38 OC Micro.mp. (9)

39 OC-Sensor.mp. (17)

40 OC Sensor.mp. (17)

41 OC-Hemodia.mp. (16)

42 OC Hemodia.mp. (16)

43 OC-Light.mp. (2)

44 OC Light.mp. (2)

45 Aimstep.mp. (0)

46 Magstream.mp. (10)

47 Immudia.mp. (6)

48 or/18,22-47 (103377)

49 exp “predictive value of tests”/ (125700)

50 exp “Sensitivity and specificity”/ (371895)

51 exp False Negative Reactions/ (15236)

52 exp False Positive Reactions/ (23021)

53 exp Reproducibility of Results/ (242180)

54 exp Reference Values/ (138214)

55 exp Diagnostic Errors/ (87977)

56 exp Reference Standards/ (31448)

57 exp Observer Variation/ (28598)

58 exp Quality Assurance, Health Care/ (236607)

59 standards.fs. (508061)

60 sensitiv$.ti,ab. (882592)

61 specificit$.ti,ab. (312517)

62 predictive value.ti,ab. (50699)

63 accurac$.ti,ab. (200813)

64 false positive$.ti,ab. (39039)

65 false negative$.ti,ab. (22776)

66 miss rate$.ti,ab. (203)

67 error rate$.ti,ab. (7478)

68 detection rate$.ti,ab. (11682)

69 diagnostic yield$.ti,ab. (4734)

70 likelihood ratio$.ti,ab. (7576)

71 odds ratio/and di.fs. (7310)

72 diagnostic odds ratio$.ti,ab. (443)

73 or/49-72 (2331963)

74 48 and 73 (23874)

75 exp Colorectal Neoplasms/ (134899)

76 exp Colonic Neoplasms/ (63388)

77 exp Sigmoid Neoplasms/ (3740)

78 exp Rectal Neoplasms/ (35360)

79 exp Intestinal Polyps/ (11351)

80 exp Colonic Polyps/ (5666)

81 colon cancer.ti,ab,hw. (26207)

82 colonic cancer.ti,ab,hw. (2114)

83 colorectal cancer.ti,ab,hw. (48264)

84 colon neoplasm.ti,ab,hw. (35)

85 colonic neoplasm.ti,ab,hw. (130)

86 colorectal neoplasm.ti,ab,hw. (220)

87 adenoma$.ti,ab,hw. (86623)

88 colon polyp.ti,ab,hw. (157)

89 colonic polyp.ti,ab,hw. (238)

90 colorectal polyp.ti,ab,hw. (181)

91 or/75-90 (223731)

92 74 and 91 (1814)

93 limit 92 to english language (1593)

94 limit 93 to humans (1532)

95 limit 94 to yr=”2008 - 2012” (447)

DARE, Cochrane Database, and HTA Search Strategy

#1 occult blood:kw

#2 “fecal occult blood”:ti,ab

#3 “faecal occult blood”:ti,ab

#4 “fecal immunochemical”:ti,ab

#5 “faecal immunochemical”:ti,ab

#6 fobt*:ti,ab

#7 ifobt*:ti,ab

#8 “instant-view”:ti,ab,kw

#9 “instant view”:ti,ab,kw

#10 “Hemoccult':ti,ab,kw

#11 “immocare”:ti,ab,kw

#12 “flexsure obt”:ti,ab,kw

#13 “monohaem”:ti,ab,kw

#14 “hemosure”:ti,ab,kw

#15 “hemoccult ict':ti,ab,kw

#16 “Occultech”:ti,ab,kw

#17 “Quickvue”:ti,ab,kw

#18 “clearview”:ti,ab,kw

#19 “hemoquant':ti,ab,kw

#20 “hema screen”:ti,ab,kw

#21 “hemascreen”:ti,ab,kw

#22 “hema-screen”:ti,ab,kw

#23 “innovacon”:ti,ab,kw

#24 “oc micro”:ti,ab,kw

#25 “oc-micro”:ti,ab,kw

#26 “oc sensor”:ti,ab,kw

#27 “oc-sensor”:ti,ab,kw

#28 “oc light”:ti,ab,kw

#29 “oc-light”:ti,ab,kw

#30 “oc hemodia”:ti,ab,kw

#31 “oc-hemodia”:ti,ab,kw

#32 “aimstep”:ti,ab,kw

#33 “magstream”:ti,ab,kw

#34 “immudia”:ti,ab,kw

#35 #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21 or #22 or #23 or #24 or #25 or #26 or #27 or #28 or #29 or #30 or #31 or #32 or #33 or #34

#36 “sensitivity”:kw

#37 “specificity”:kw

#38 “predictive value”:kw

#39 “roc”:kw

#40 receiver next operat*:kw

#41 false next negative:kw

#42 false next positive:kw

#43 diagnostic next error*:kw

#44 reproducibility:kw

#45 reference next value*:kw

#46 refererence next standards:kw

#47 diagnostic next accuracy:kw

#48 diagnostic next value:kw

#49 #36 or #37 or #38 or #39 or #40 or #41 or #42 or #43 or #44 or #45 or #46 or #47 or #48

#50 #35 and #49

#51 sensitiv*.ti,ab

#52 specificit*:ti,ab

#53 predictive next value:ti,ab

#54 accurac*:ti,ab

#55 miss next rate*:ti,ab

#56 detection next rate*:ti,ab

#57 diagnostic next yield*:ti,ab

#58 likelihood next ratio*.ti,ab

#59 diagnostic next odds next ratio*.ti,ab

#60 “odds ratio” and “diagnosis”:kw

#61 #52 or #53 or #54 or #55 or #56 or #57 or #58 or #59 or #60

#62 #35 and #61

#63 #50 or #62

#64 screening:kw

#65 screen*:ti,ab

#66 #64 or #65

#67 #63 and #66 from 2008 to 2012

EMBASE Search Strategy

#1 Fecal

#2 faecal

#3 feces

#4 #1 OR #2 OR #3

#5 ‘immunochemistry’/exp

#6 immunochem*

#7 #5 OR #6

#8 #4 AND #7

#9 fit:ab,ti

#10 guaiac:ab,ti

#11 ‘occult blood’

#12 fobt*

#13 fob*

#14 ifobt

#15 ifob*

#16 #6 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15

#17 insure

#18 inform

#19 #17 OR #18

#20 #16 AND #19

#21 ‘instant view’

#22 hemoccult

#23 immocare

#24 flexsure

#25 monohaem

#26 hemosure

#27 occultech

#28 quickvue

#29 clearview

#30 hemoquant

#31 ‘hema screen’

#32 innovacon

#33 ‘oc micro’

#34 ‘oc sensor’

#35 ‘oc hemodia’

#36 ‘oc light’

#37 aimstep

#38 magstream

#39 immudia

#40 #16 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29 OR #30 OR #31 OR #32 OR #33 OR #34 OR #35 OR #36 OR #37 OR #38 OR #39

#41 ‘predictive value’/de

#42 ‘sensitivity and specificity’/de

#43 ‘laboratory diagnosis’/exp

#44 ‘reproducibility’/de

#45 ‘reference value’/de

#46 ‘diagnostic error’/exp

#47 ‘diagnostic test accuracy study’/de

#48 ‘diagnostic accuracy’/de

#49 ‘diagnostic value’/de

#50 ‘standard’/de

#51 ‘gold standard’/de

#52 ‘observer variation’/de

#53 ‘health care quality’/de

#54 ‘biomedical technology assessment’/de

#55 ‘clinical effectiveness’/de

#56 ‘clinical indicator’/de

#57 ‘medical error’/exp

#58 ‘root cause analysis’/de

#59 ‘good laboratory practice’/de

#60 ‘validation process’/de

#61 sensitiv*:ab,ti

#62 specificit*:ab,ti

#63 ‘predictive value’:ab,ti

#64 accurac*:ab,ti

#65 (false NEXT/1 positive*):ab,ti

#66 (false NEXT/1 negative*):ab,ti

#67 (miss NEXT/1 rate*):ab,ti

#68 (error NEXT/1 rate*):ab,ti

#69 (detection NEXT/1 rate*):ab,ti

#70 (diagnostic NEXT/1 yield*):ab,ti

#71 (likelihood NEXT/1 ratio*):ab,ti

#72 ‘odds ratio’:ab,ti AND diagnosis:ab,ti

#73 risk:ab,ti AND diagnosis:ab,ti

#74 ‘diagnostic odds ratio’:ab,ti OR‘diagnostic odds ratios’:ab,ti

#75 ‘diagnostic accuracy’

#76 ‘reference standard’:ab,ti OR ‘reference standards’:ab,ti

#77 #41 OR #42 OR #43 OR #44 OR #45 OR #46 OR #47 OR #48 OR #49 OR #50 OR #51 OR #52 OR #53 OR #54 OR #55 OR #56 OR #57 OR #58 OR #59 OR #60 OR #61 OR #62 OR #63 OR #64 OR #65 OR #66 OR #67 OR #68 OR #69 OR #70 OR #71 OR #72 OR #73 OR #74 OR #75 OR #76

#78 #40 AND #77

#79 ‘colon tumor’/exp

#80 ‘rectum tumor’/exp

#81 ‘intestine polyp’/exp

#82 ‘colon polyp’/exp

#83 ‘colon cancer’:ab,ti

#84 ‘colonic cancer’:ab,ti

#85 ‘colorectal cancer’:ab,ti

#86 ‘colon neoplasm’:ab,ti

#87 ‘colonic neoplasm’:ab,ti

#88 ‘colorectal neoplasm’:ab,ti

#89 adenoma*:ab,ti

#90 ‘colon polyp’:ab,ti

#91 ‘colonic polyp’:ab,ti

#92 ‘colorectal polyp’:ab,ti

#93 #79 OR #80 OR #81 OR #82 OR #83 OR #84 OR #85 OR #86 OR #87 OR #88 OR #89 OR #90 OR #91 OR #92

#94 #78 AND #93

#95 #78 AND #93 AND [english]/lim

#96 #78 AND #93 AND [english]/lim NOT ([animals]/lim NOT ([humans]/lim OR 'patient'/exp))

#97 #78 AND #93 AND [english]/lim NOT ([animals]/lim NOT ([humans]/lim OR 'patient'/exp)) NOT 'conference abstract'/it

#98 #78 AND #93 AND [english]/lim NOT ([animals]/lim NOT ([humans]/lim OR 'patient'/exp)) NOT 'conference abstract'/it AND [2008-2012]/py

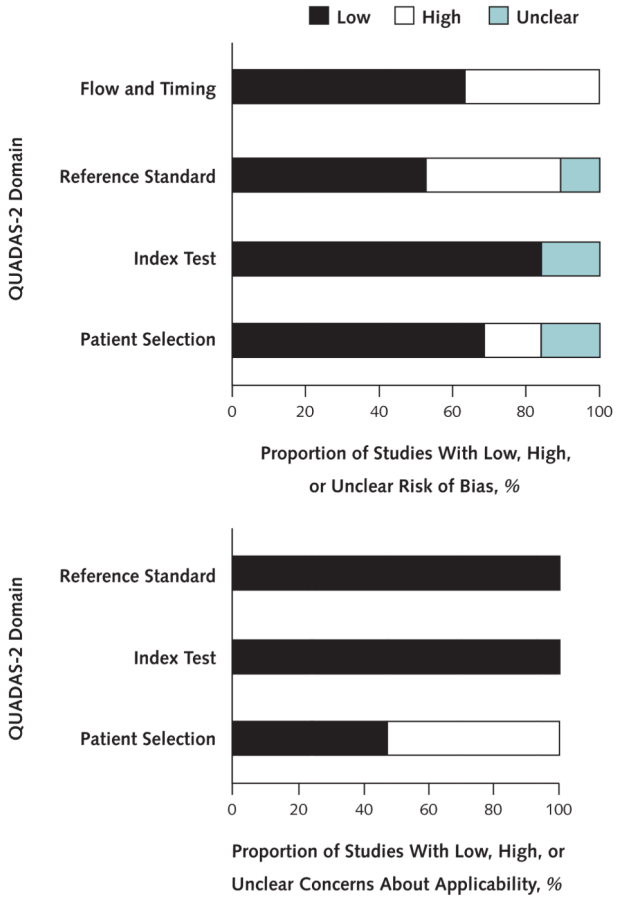
Appendix Table 2
Quality Assessment of Diagnostic Accuracy Studies 2
Results for Included Studies

Author, Year (Reference)	Risk of Bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Re
Sohn et al, 2005 (14)	Unclear risk	Unclear risk	Unclear risk	Low risk	High risk	Low risk	Lo
Levi et al, 2011 (15)	Low risk	Low risk	High risk	High risk	Low risk	Low risk	Lo
Allison et al, 1996 (31)	High risk	Low risk	High risk	High risk	Low risk	Low risk	Lo
Allison et al, 2007 (32)	High risk	Low risk	High risk	High risk	High risk	Low risk	Lo
Levi et al, 2007 (33)	Low risk	Low risk	Low risk	Low risk	High risk	Low risk	Lo
Cheng et al, 2002 (34)	High risk	Low risk	Low risk	Low risk	High risk	Low risk	Lo
Morikawa et al, 2005 (35)	Low risk	Low risk	Low risk	Low risk	High risk	Low risk	Lo
Nakama et al, 1999 (36)	Unclear risk	Unclear risk	Unclear risk	Low risk	High risk	Low risk	Lo
Nakama et al, 1996 (37)	Unclear risk	Unclear risk	High risk	High risk	High risk	Low risk	Lo
Launoy et al, 2005 (38)	Low risk	Low risk	High risk	High risk	Low risk	Low risk	Lo
Itoh et al, 1996 (39)	Low risk	Low risk	High risk	High risk	Low risk	Low risk	Lo
Nakazato et al, 2006 (40)	Low risk	Low risk	Low risk	Low risk	High risk	Low risk	Lo
Park et al, 2010 (41)	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Lo
de Wijkerslooth et al, 2012 (42)	Low risk	Low risk	Low risk	Low risk	High risk	Low risk	Lo
Parra-Blanco et al, 2010 (43)	Low risk	Low risk	High risk	High risk	Low risk	Low risk	Lo
Chiu et al, 2013 (44)	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Lo
Chiang et al, 2011 (45)	Low risk	Low risk	Low risk	Low risk	High risk	Low risk	Lo
Brenner and Tao, 2013 (46)	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Lo
Brenner and Tao, 2013 (46)	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Lo

Appendix Table 3
Results of Bivariate Meta-regression With Covariates

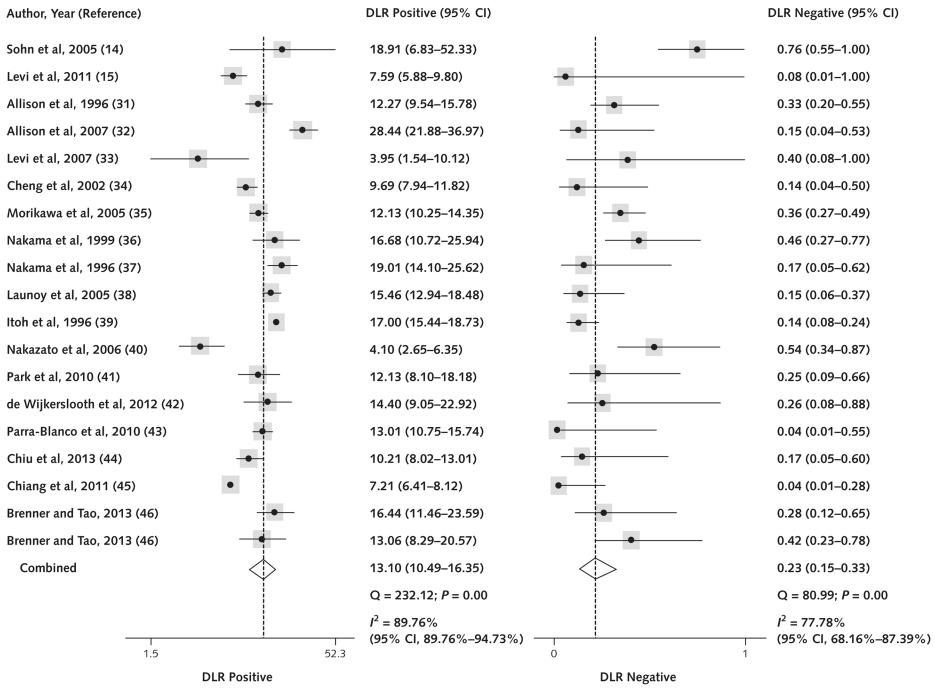
Category	Studies, *n*	Sensitivity (95% CI)	*P* Value	Specificity (95% CI)	*P* Value
Geographic region					
Asian countries	10	0.75 (0.63–0.87)	0.04	0.94 (0.92–0.96)	<0.01
Non-Asian countries	9	0.83 (0.72–0.94)	0.04	0.94 (0.92–0.96)	<0.01
FIT format					
Quantitative	11	0.73 (0.61–0.85)	0.01	0.94 (0.92–0.96)	<0.01
Qualitative	8	0.85 (0.75–0.95)	0.01	0.94 (0.91–0.96)	<0.01
Patient age					
<40 y	4	0.74 (0.55–0.94)	0.24	0.94 (0.91–0.97)	<0.01
≥40 y	14	0.79 (0.70–0.89)	0.24	0.94 (0.93–0.96)	<0.01

FIT = fecal immunochemical test.



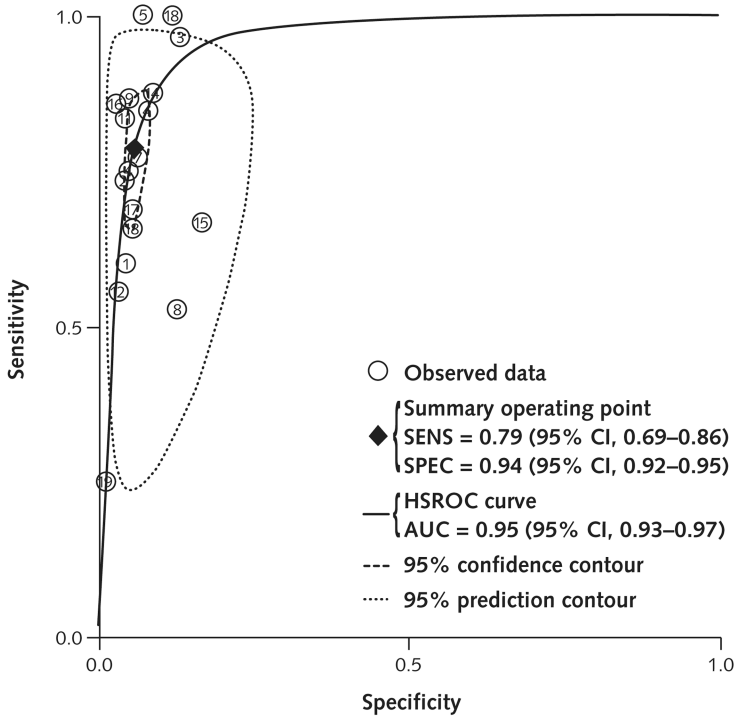
Appendix Figure 1. Overall quality assessment of included studies

QUADAS-2 = Quality Assessment of Diagnostic Accuracy Studies 2.



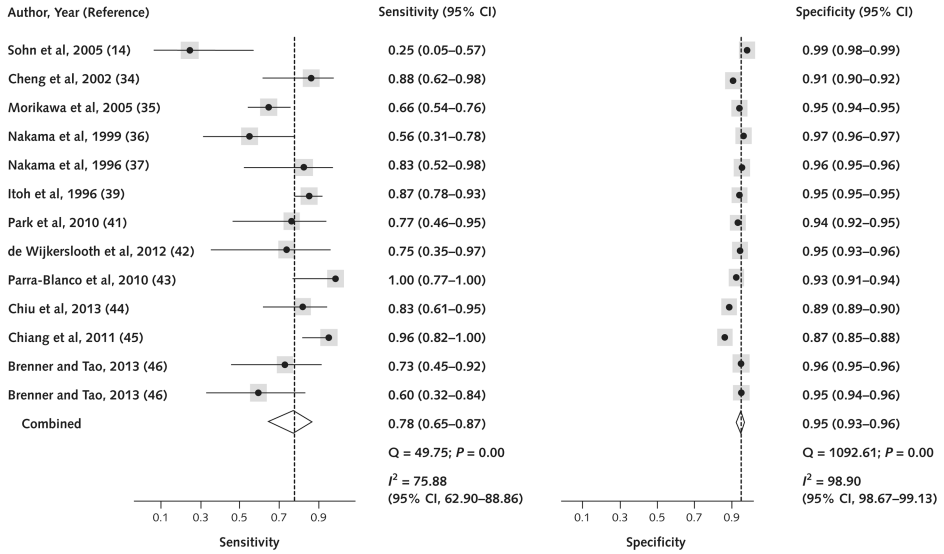
Appendix Figure 2. Pooled positive and negative likelihood ratios for fecal immunochemical tests for the detection of colorectal cancer for all included studies

The circles in squares represent the point estimate, the horizontal lines represent the 95% CI, the dotted lines represent the pooled estimate, and the diamonds represent the 95% CI of the pooled estimate. DLR = diagnostic likelihood ratio.



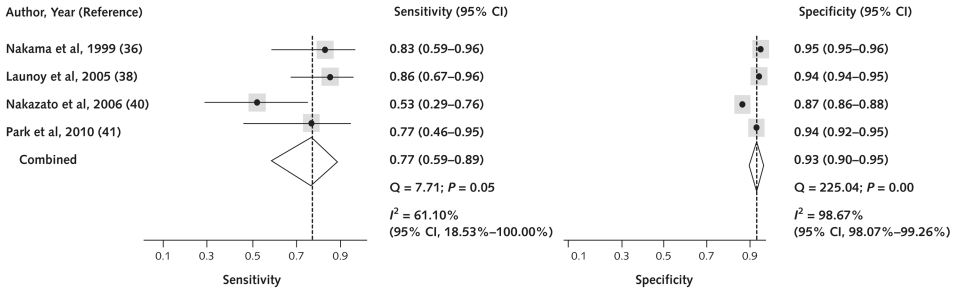
Appendix Figure 3. HSROC curve of the sensitivity versus specificity of fecal immunochemical tests for the detection of colorectal cancer for all included studies

The circles represent the data from each included study, the straight line represents the curve, and the diamond represents the summary point of the curve to which the pooled sensitivity and specificity correspond. The dashed line represents the 95% confidence area for the summary point, and the dotted line represents the 95% confidence area in which a new diagnostic accuracy fecal immunochemical test study will be located. AUC = area under the curve; SENS = sensitivity; SPEC = specificity; HSROC = hierarchical summary receiver-operating characteristic.



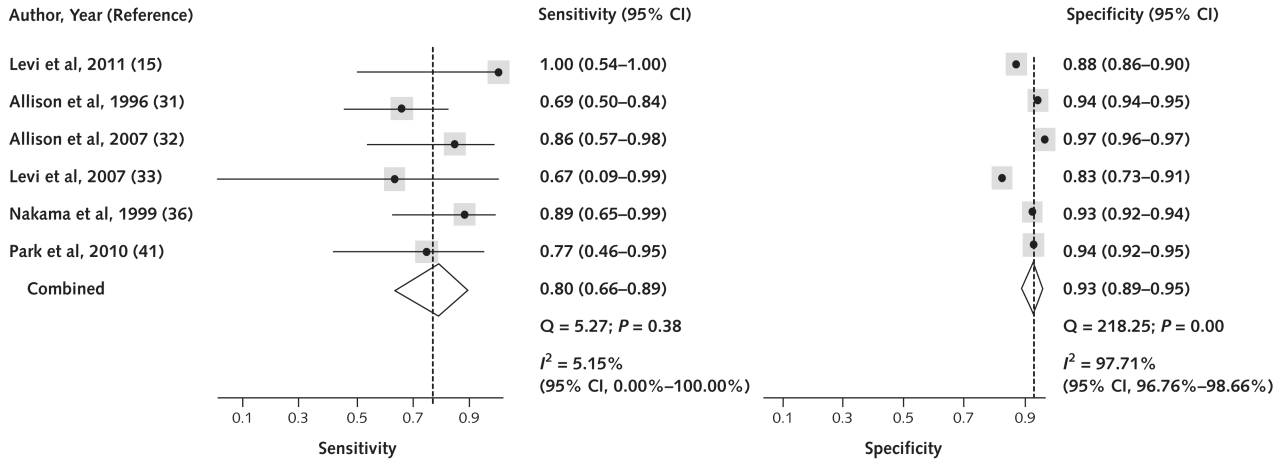
Appendix Figure 4. Subgroup analysis: pooled sensitivity and specificity for 1-sample fecal immunochemical test studies

The circles in squares represent the point estimate, the horizontal lines represent the 95% CI, the dotted lines represent the pooled estimate, and the diamonds represent the 95% CI of the pooled estimate.



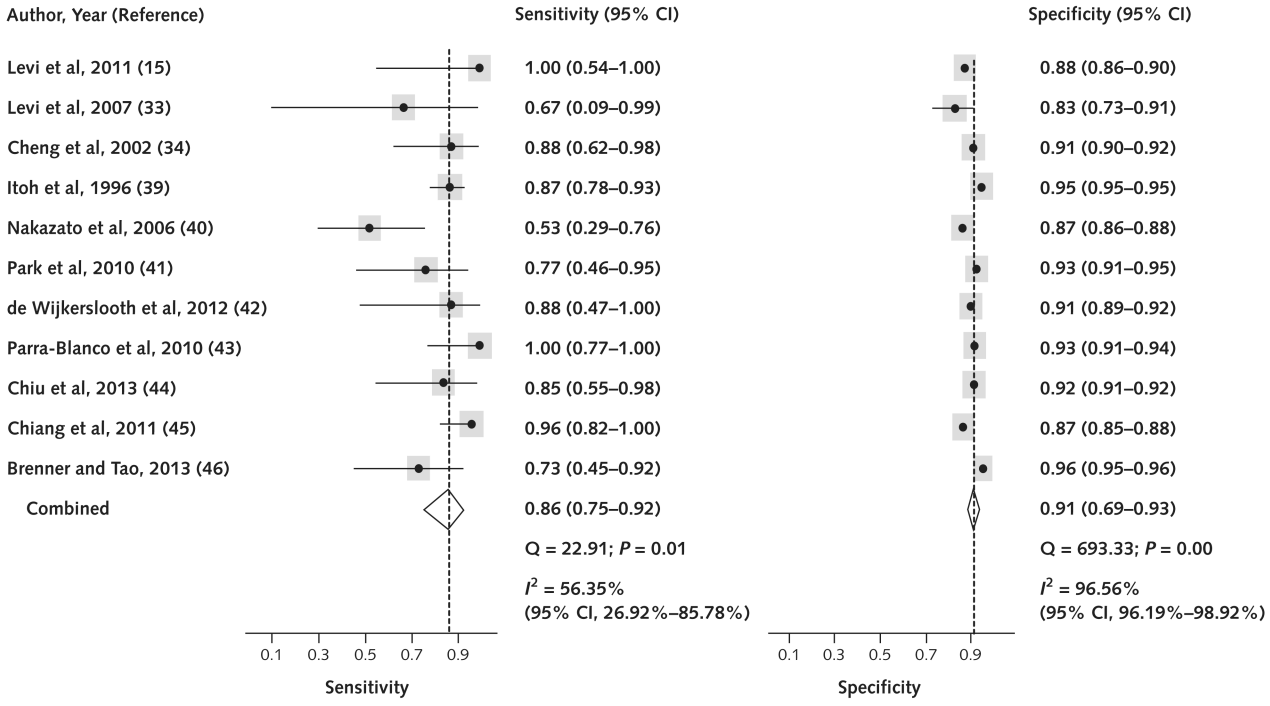
Appendix Figure 5. Subgroup analysis: pooled sensitivity and specificity for 2-sample fecal immunochemical test studies

The circles in squares represent the point estimate, the horizontal lines represent the 95% CI, the dotted lines represent the pooled estimate, and the diamonds represent the 95% CI of the pooled estimate.



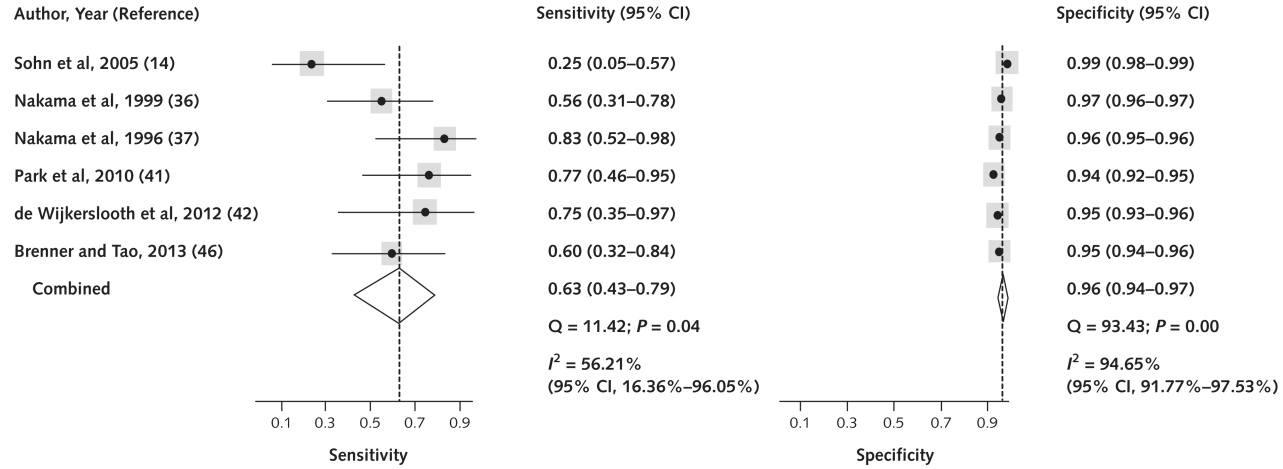
Appendix Figure 6. Subgroup analysis: pooled sensitivity and specificity for 3-sample fecal immunochemical test studies

The circles in squares represent the point estimate, the horizontal lines represent the 95% CI, the dotted lines represent the pooled estimate, and the diamonds represent the 95% CI of the pooled estimate.



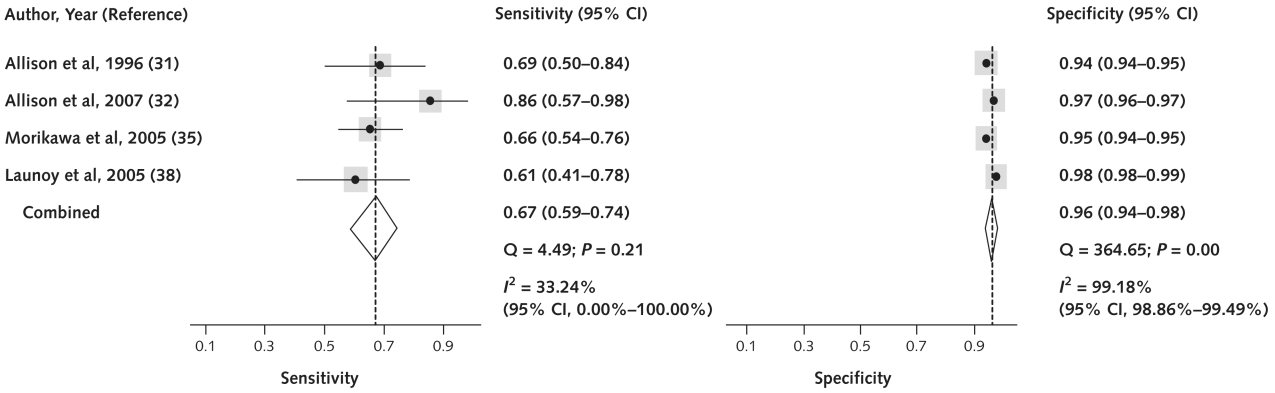
Appendix Figure 7. Subgroup analysis: pooled sensitivity and specificity for fecal immunochemical test studies with cutoff <20 μg/g

The circles in squares represent the point estimate, the horizontal lines represent the 95% CI, the dotted lines represent the pooled estimate, and the diamonds represent the 95% CI of the pooled estimate.



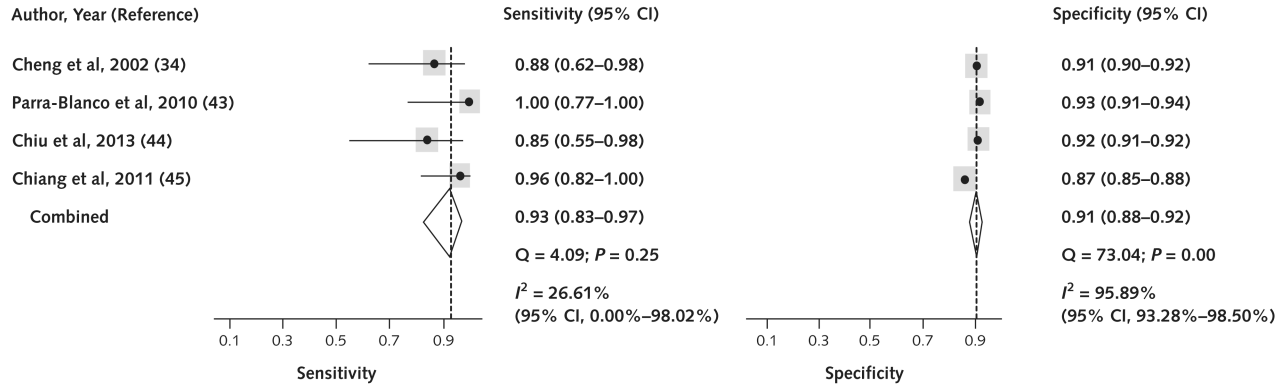
Appendix Figure 8. Subgroup analysis: pooled sensitivity and specificity for fecal immunochemical test studies with cutoff 20–50 μg/g

The circles in squares represent the point estimate, the horizontal lines represent the 95% CI, the dotted lines represent the pooled estimate, and the diamonds represent the 95% CI of the pooled estimate.



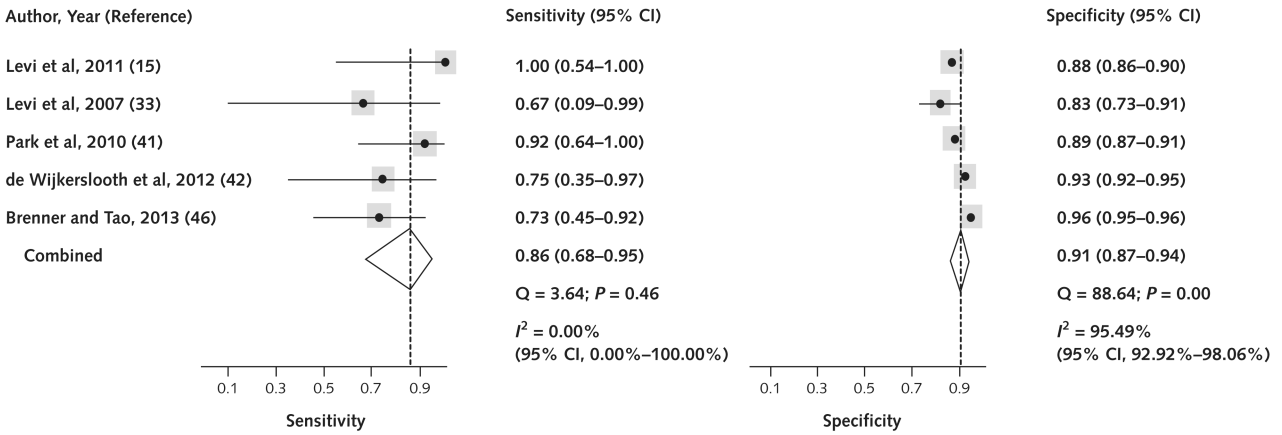
Appendix Figure 9. Subgroup analysis: pooled sensitivity and specificity for fecal immunochemical test studies with cutoff >50 μg/g

The circles in squares represent the point estimate, the horizontal lines represent the 95% CI, the dotted lines represent the pooled estimate, and the diamonds represent the 95% CI of the pooled estimate.



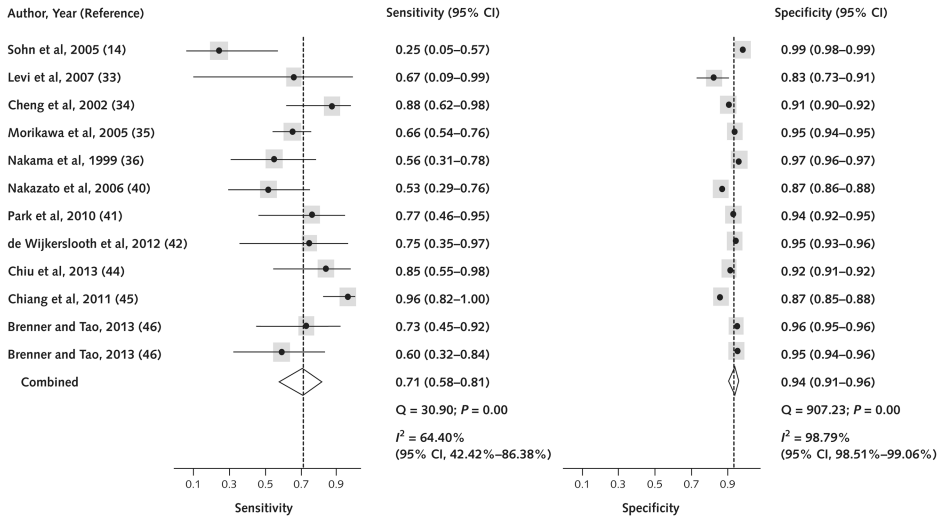
Appendix Figure 10. Subgroup analysis: pooled sensitivity and specificity for OC-Light fecal immunochemical tests

The circles in squares represent the point estimate, the horizontal lines represent the 95% CI, the dotted lines represent the pooled estimate, and the diamonds represent the 95% CI of the pooled estimate.



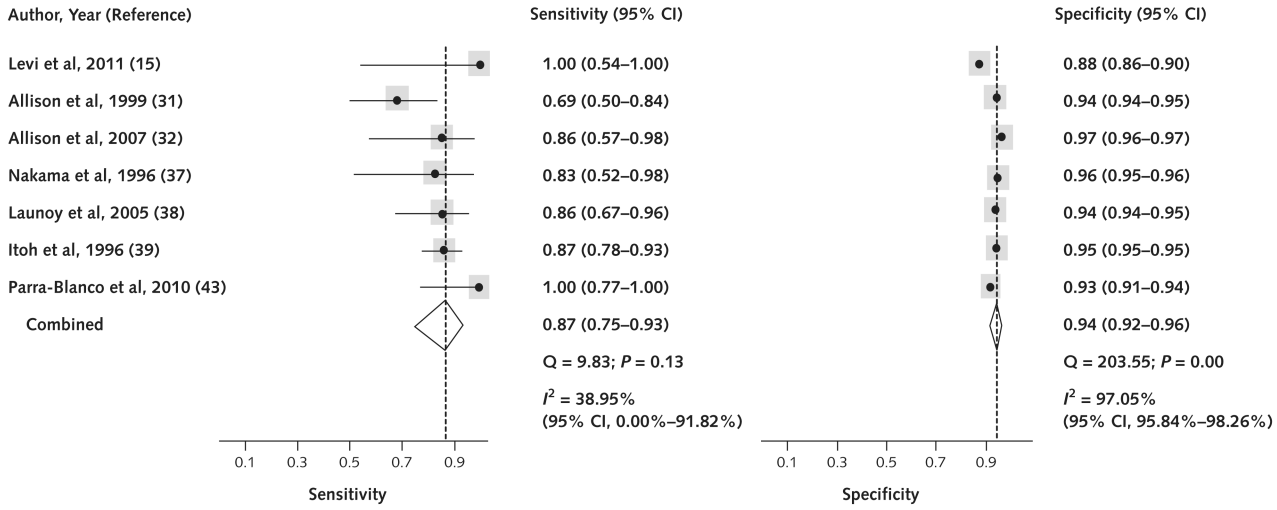
Appendix Figure 11. Subgroup analysis: pooled sensitivity and specificity for OC-Micro fecal immunochemical tests

The circles in squares represent the point estimate, the horizontal lines represent the 95% CI, the dotted lines represent the pooled estimate, and the diamonds represent the 95% CI of the pooled estimate.



Appendix Figure 12. Subgroup analysis: pooled sensitivity and specificity for studies using colonoscopy to follow-up of patients with negative fecal immunochemical test results

The circles in squares represent the point estimate, the horizontal lines represent the 95% CI, the dotted lines represent the pooled estimate, and the diamonds represent the 95% CI of the pooled estimate.



Appendix Figure 13. Subgroup analysis: pooled sensitivity and specificity for studies using a 2-y longitudinal follow-up of patients with negative fecal immunochemical test results

The circles in squares represent the point estimate, the horizontal lines represent the 95% CI, the dotted lines represent the pooled estimate, and the diamonds represent the 95% CI of the pooled estimate.

References

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. CA Cancer J Clin. 2013; 63:11–30. [PubMed: 23335087]

2. Hardcastle JD, Chamberlain JO, Robinson MH, Moss SM, Amar SS, Balfour TW, et al. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. Lancet. 1996; 348:1472–7. [PubMed: 8942775]

3. Kronborg O, Fenger C, Olsen J, Jørgensen OD, Søndergaard O. Randomised study of screening for colorectal cancer with faecal-occult-blood test. Lancet. 1996; 348:1467–71. [PubMed: 8942774]

4. Mandel JS, Bond JH, Church TR, Snover DC, Bradley GM, Schuman LM, et al. Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota Colon Cancer Control Study. N Engl J Med. 1993; 328:1365–71. [PubMed: 8474513]

5. Lieberman DA, Weiss DG, Veterans Affairs Cooperative Study Group 380. One-time screening for colorectal cancer with combined fecal occult-blood testing and examination of the distal colon. N Engl J Med. 2001; 345:555–60. [PubMed: 11529208]

6. Imperiale TF, Ransohoff DF, Itzkowitz SH, Turnbull BA, Ross ME, Colorectal Cancer Study Group. Fecal DNA versus fecal occult blood for colorectal-cancer screening in an average-risk population. N Engl J Med. 2004; 351:2704–14. [PubMed: 15616205]

7. Gellad ZF, Stechuchak KM, Fisher DA, Olsen MK, McDuffie JR, Ostbye T, et al. Longitudinal adherence to fecal occult blood testing impacts colorectal cancer screening quality. Am J Gastroenterol. 2011; 106:1125–34. [PubMed: 21304501]

8. Fenton JJ, Elmore JG, Buist DS, Reid RJ, Tancredi DJ, Baldwin LM. Longitudinal adherence with fecal occult blood test screening in community practice. Ann Fam Med. 2010; 8:397–401. [PubMed: 20843880]

9. Rabeneck L, Rumble RB, Thompson F, Mills M, Oleschuk C, Whibley A, et al. Fecal immunochemical tests compared with guaiac fecal occult blood tests for population-based colorectal cancer screening. Can J Gastroenterol. 2012; 26:131–47. [PubMed: 22408764]

10. Whitlock EP, Lin JS, Liles E, Beil TL, Fu R. Screening for colorectal cancer: a targeted, updated systematic review for the U.S. Preventive Services Task Force. Ann Intern Med. 2008; 149:638–58. [PubMed: 18838718]

11. Levin B, Lieberman DA, McFarland B, Andrews KS, Brooks D, Bond J, et al. American Cancer Society Colorectal Cancer Advisory Group. Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. Gastroenterology. 2008; 134:1570–95. [PubMed: 18384785]

12. Chen LS, Yen AM, Chiu SY, Liao CS, Chen HH. Baseline faecal occult blood concentration as a predictor of incident colorectal neoplasia: longitudinal follow-up of a Taiwanese population-based colorectal cancer screening cohort. Lancet Oncol. 2011; 12:551–8. [PubMed: 21592859]

13. Stegeman I, de Wijkerslooth TR, Mallant-Hent RC, de Groot K, Stroobants AK, Fockens P, et al. Implementation of population screening for colorectal cancer by repeated Fecal Immunochemical Test (FIT): third round. BMC Gastroenterol. 2012; 12:73. [PubMed: 22713100]

14. Sohn DK, Jeong SY, Choi HS, Lim SB, Huh JM, Kim DH, et al. Single immunochemical fecal occult blood test for detection of colorectal neoplasia. Cancer Res Treat. 2005; 37:20–3. [PubMed: 19956505]

15. Levi Z, Birkenfeld S, Vilkin A, Bar-Chana M, Lifshitz I, Chared M, 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. [PubMed: 20658527]

16. Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM, Cochrane Diagnostic Test Accuracy Working Group. Systematic reviews of diagnostic test accuracy. Ann Intern Med. 2008; 149:889–97. [PubMed: 19075208]

17. Pai M, McCulloch M, Enanoria W, Colford JM Jr. Systematic reviews of diagnostic test evaluations: what's behind the scenes? [Editorial]. ACP J Club. 2004; 141:A11–3. [PubMed: 15230574]

18. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Standards for Reporting of Diagnostic Accuracy. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. Ann Intern Med. 2003; 138:40–4. [PubMed: 12513043]

19. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg. 2010; 8:336–41. [PubMed: 20171303]
20. Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. CMAJ. 2006; 174:469–76. [PubMed: 16477057]
21. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011; 155:529–36. [PubMed: 22007046]
22. Fraser CG, Allison JE, Halloran SP, Young GP, Expert Working Group on Fecal Immunochemical Tests for Hemoglobin, Colorectal Cancer Screening Committee, World Endoscopy Organization. A proposal to standardize reporting units for fecal immunochemical tests for hemoglobin. J Natl Cancer Inst. 2012; 104:810–4. [PubMed: 22472305]
23. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. JAMA. 1994; 271:703–7. [PubMed: 8309035]
24. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol. 2005; 58:982–90. [PubMed: 16168343]
25. Zwinderman AH, Bossuyt PM. We should not pool diagnostic likelihood ratios in systematic reviews. Stat Med. 2008; 27:687–97. [PubMed: 17611957]
26. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med. 2001; 20:2865–84. [PubMed: 11568945]
27. Swets JA. Measuring the accuracy of diagnostic systems. Science. 1988; 240:1285–93. [PubMed: 3287615]
28. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003; 327:557–60. [PubMed: 12958120]
29. Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA. A unification of models for meta-analysis of diagnostic accuracy studies. Biostatistics. 2007; 8:239–51. [PubMed: 16698768]
30. Dwamena, Ben A. MIDAS: Statamodule for meta-analytical integration of diagnostic accuracy studies. 2007. Accessed on http://econpapers.repec.org/software/bocbocode/s456880.htm on 14 August 2013
31. Allison JE, Tekawa IS, Ransom LJ, Adrain AL. A comparison of fecal occult-blood tests for colorectal-cancer screening. N Engl J Med. 1996; 334:155–9. [PubMed: 8531970]
32. Allison JE, Sakoda LC, Levin TR, Tucker JP, Tekawa IS, Cuff T, et al. Screening for colorectal neoplasms with new fecal occult blood tests: update on performance characteristics. J Natl Cancer Inst. 2007; 99:1462–70. [PubMed: 17895475]
33. Levi Z, Rozen P, Hazazi R, Vilkin A, Waked A, Maoz E, et al. A quantitative immunochemical fecal occult blood test for colorectal neoplasia. Ann Intern Med. 2007; 146:244–55. [PubMed: 17310048]
34. Cheng TI, Wong JM, Hong CF, Cheng SH, Cheng TJ, Shieh MJ, et al. Colorectal cancer screening in asymptomaic adults: comparison of colonoscopy, sigmoidoscopy and fecal occult blood tests. J Formos Med Assoc. 2002; 101:685–90. [PubMed: 12517041]
35. Morikawa T, Kato J, Yamaji Y, Wada R, Mitsushima T, Shiratori Y. A comparison of the immunochemical fecal occult blood test and total colonoscopy in the asymptomatic population. Gastroenterology. 2005; 129:422–8. [PubMed: 16083699]
36. Nakama H, Yamamoto M, Kamijo N, Li T, Wei N, Fattah AS, et al. Colonoscopic evaluation of immunochemical fecal occult blood test for detection of colorectal neoplasia. Hepatogastroenterology. 1999; 46:228–31. [PubMed: 10228797]
37. Nakama H, Kamijo N, Abdul Fattah AS, Zhang B. Validity of immunological faecal occult blood screening for colorectal cancer: a follow up study. J Med Screen. 1996; 3:63–5. [PubMed: 8849761]

38. Launoy GD, Bertrand HJ, Berchi C, Talbourdet VY, Guizard AV, Bouvier VM, 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. [PubMed: 15700317]
39. Itoh M, Takahashi K, Nishida H, Sakagami K, Okubo T. 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. [PubMed: 8849762]
40. Nakazato M, Yamano HO, Matsushita HO, Sato K, Fujita K, Yamanaka Y, et al. Immunologic fecal occult blood test for colorectal cancer screening. Japan Medical Association Journal. 2006; 49:203–207.
41. Park DI, Ryu S, Kim YH, Lee SH, Lee CK, Eun CS, et al. Comparison of guaiac-based and quantitative immunochemical fecal occult blood testing in a population at average risk undergoing colorectal cancer screening. Am J Gastroenterol. 2010; 105:2017–25. [PubMed: 20502450]
42. de Wijkerslooth TR, Stoop EM, Bossuyt PM, Meijer GA, van Ballegooijen M, van Roon AH, et al. Immunochemical fecal occult blood testing is equally sensitive for proximal and distal advanced neoplasia. Am J Gastroenterol. 2012; 107:1570–8. [PubMed: 22850431]
43. Parra-Blanco A, Gimeno-García AZ, Quintero E, Nicolás D, Moreno SG, Jiménez A, et al. Diagnostic accuracy of immunochemical versus guaiac faecal occult blood tests for colorectal cancer screening. J Gastroenterol. 2010; 45:703–12. [PubMed: 20157748]
44. Chiu HM, Lee YC, Tu CH, Chen CC, Tseng PH, Liang JT, et al. Association between early stage colon neoplasms and false-negative results from the fecal immunochemical test. Clin Gastroenterol Hepatol. 2013; 11:832–8.e1-2. [PubMed: 23376002]
45. Chiang TH, Lee YC, Tu CH, Chiu HM, Wu MS. Performance of the immunochemical fecal occult blood test in predicting lesions in the lower gastrointestinal tract. CMAJ. 2011; 183:1474–81. [PubMed: 21810951]
46. Brenner H, Tao S. Superior diagnostic performance of faecal immunochemical tests for haemoglobin in a head-to-head comparison with guaiac based faecal occult blood test among 2235 participants of screening colonoscopy. Eur J Cancer. 2013; 49:3049–54. [PubMed: 23706981]
47. Bressler B, Paszat LF, Chen Z, Rothwell DM, Vinden C, Rabeneck L. Rates of new or missed colorectal cancers after colonoscopy and their risk factors: a population-based analysis. Gastroenterology. 2007; 132:96–102. [PubMed: 17241863]
48. Hosokawa O, Shirasaki S, Kaizaki Y, Hayashi H, Douden K, Hattori M. Invasive colorectal cancer detected up to 3 years after a colonoscopy negative for cancer. Endoscopy. 2003; 35:506–10. [PubMed: 12783349]
49. Avidan B, Sonnenberg A, Schnell TG, Leya J, Metz A, Sontag SJ. New occurrence and recurrence of neoplasms within 5 years of a screening colonoscopy. Am J Gastroenterol. 2002; 97:1524–9. [PubMed: 12094877]
50. Sawhney MS, Farrar WD, Gudiseva S, Nelson DB, Lederle FA, Rector TS, et al. Microsatellite instability in interval colon cancers. Gastroenterology. 2006; 131:1700–5. [PubMed: 17087932]
51. Rex DK, Rahmani EY, Haseman JH, Lemmel GT, Kaster S, Buckley JS. Relative sensitivity of colonoscopy and barium enema for detection of colorectal cancer in clinical practice. Gastroenterology. 1997; 112:17–23. [PubMed: 8978337]
52. Farrar WD, Sawhney MS, Nelson DB, Lederle FA, Bond JH. Colorectal cancers found after a complete colonoscopy. Clin Gastroenterol Hepatol. 2006; 4:1259–64. [PubMed: 16996804]
53. Pickhardt PJ, Choi JR, Hwang I, Butler JA, Puckett ML, Hildebrandt HA, et al. Computed tomographic virtual colonoscopy to screen for colorectal neoplasia in asymptomatic adults. N Engl J Med. 2003; 349:2191–200. [PubMed: 14657426]
54. Morrison A, Polisena J, Husereau D, Moulton K, Clark M, Fiander M, et al. The effect of English-language restriction on systematic review-based meta-analyses: a systematic review of empirical studies. Int J Technol Assess Health Care. 2012; 28:138–44. [PubMed: 22559755]
55. Pham B, Klassen TP, Lawson ML, Moher D. Language of publication restrictions in systematic reviews gave different results depending on whether the intervention was conventional or complementary. J Clin Epidemiol. 2005; 58:769–76. [PubMed: 16086467]

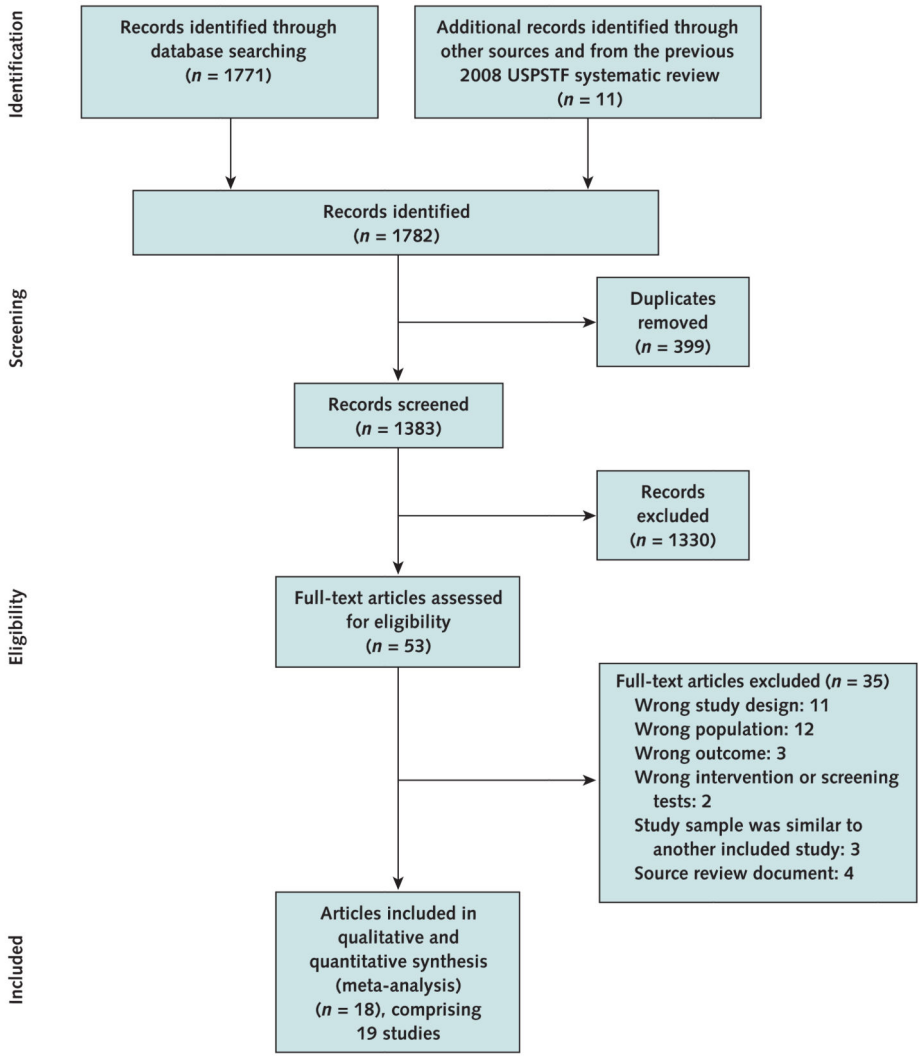


Figure 1. Summary of evidence search and selection

USPSTF = U.S. Preventive Services Task Force.

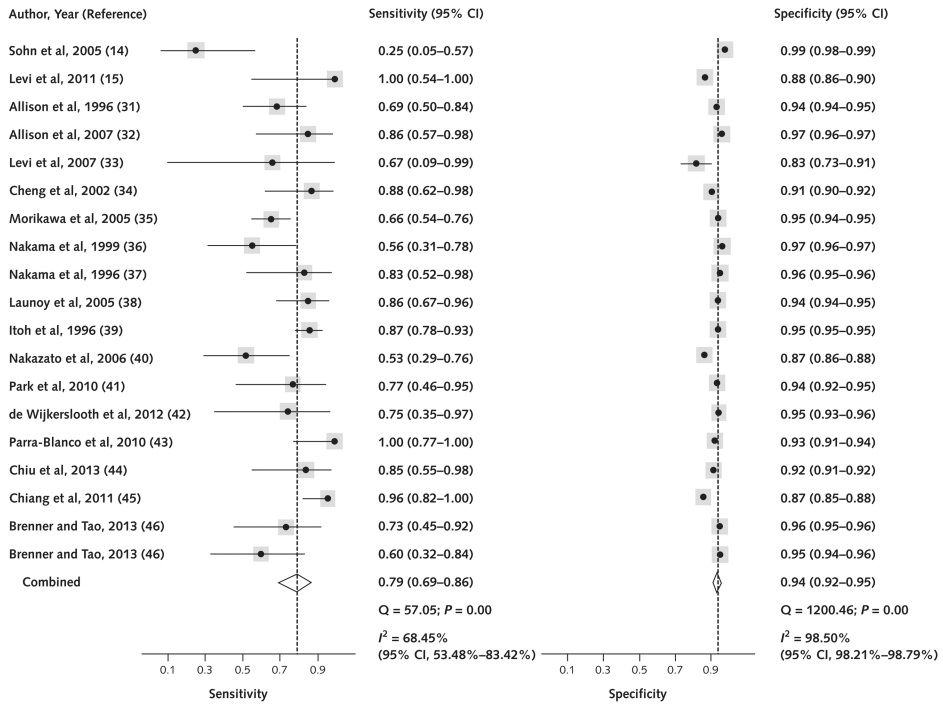


Figure 2. Pooled sensitivity and specificity for fecal immunochemical tests for the detection of colorectal cancer for all included studies

The circles in squares represent the point estimate, the horizontal lines represent the 95% CI, the dotted lines represent the pooled estimate, and the diamonds represent the 95% CI of the pooled estimate.

Table 1

Characteristics of Included Studies in Meta-analysis

Study, Year (Reference)	FIT Brand	Country	FIT Samples, *n*	Cutoff Value for a Positive Test Result, *μg/g*	Cohort Size, *n*	CRC Cases, *n*
Sohn et al, 2005 (14)	OC-Hemodia†	Korea	1	20	3794	12
Levi et al, 2011 (15)	OC-Micro	Israel	3	14	1204	6
Allison et al, 1996 (31)	HemeSelect†	United States	3	100	7493	35
Allison et al, 2007 (32)	FlexSure OBT	United States	3	300	5356	14
Levi et al, 2007 (33)	OC-Micro	Israel	3	15	80	3
Cheng et al, 2002 (34)	OC-Light	Taiwan	1	10	7411	16
Morikawa et al, 2005 (35)	MagStream HemSp	Japan	1	67	21 805	79
Nakama et al, 1999 (36)	Monohaem	Japan	1	20	4611	18
Nakama et al, 1996 (37)	Monohaem	Japan	1	20	3365	12
Launoy et al, 2005 (38)	MagStream HemSp	France	2	67	7421	28
Itoh et al, 1996 (39)	OC-Hemodia†	Japan	1	10	27 860	89
Nakazato et al, 2006 (40)	OC-Hemodia†	Japan	2	16	3090	19
Park et al, 2010 (41)	OC-Micro	Korea	1	20	770	13
de Wijkerslooth et al, 2012 (42)	OC-Micro	The Netherlands	1	20	1256	8
Parra-Blanco et al, 2010 (43)	OC-Light	Spain	1	10	1756	14
Chiu et al, 2013 (44)	OC-Light	Taiwan	1	10	8822	13
Chiang et al, 2011 (45)	OC-Light	Taiwan	1	10	2796	28
Brenner and Tao, 2013 (46)	OC-Micro	Germany	1	6.1	2235	15
Brenner and Tao, 2013 (46)	Ridascreen Haemoglobin	Germany	1	24.5	2235	15

Mean Age, *y*	Reference Standard*	Sensitivity (95% CI)	Specificity (95% CI)	Positive LR (95% CI)	Negative LR (95% CI)
48.9	Colonoscopy	0.25 (0.05–0.57)	0.99 (0.98–0.99)	18.91 (6.83–52.33)	0.76 (0.55–1.00)
60.4	2-y follow-up	1.00 (0.54–1.00)	0.88 (0.86–0.90)	7.59 (5.88–9.80)	0.08 (0.01–1.00)
NR‡	2-y follow-up	0.69 (0.50–0.84)	0.94 (0.94–0.95)	12.27 (9.54–15.78)	0.33 (0.20–0.55)
58.8	2-y follow-up	0.86 (0.57–0.98)	0.97 (0.96–0.97)	28.44 (21.88–36.97)	0.15 (0.04–0.53)
NR	Colonoscopy	0.67 (0.09–0.99)	0.83 (0.73–0.91)	3.95 (1.54–10.12)	0.40 (0.08–1.00)

Mean Age, *y*	Reference Standard*	Sensitivity (95% CI)	Specificity (95% CI)	Positive LR (95% CI)	Negative LR (95% CI)
46.8	Colonoscopy	0.88 (0.62–0.98)	0.91 (0.90–0.92)	9.69 (7.94–11.82)	0.14 (0.04–0.50)
48.2	Colonoscopy	0.66 (0.54–0.76)	0.95 (0.94–0.95)	12.13 (10.25–14.35)	0.36 (0.27–0.49)
NR	Colonoscopy	0.56 (0.29–0.76)	0.97 (0.96–0.97)	16.68 (10.72–25.94)	0.46 (0.27–0.77)
NR§	2-y follow-up	0.83 (0.52–0.98)	0.96 (0.95–0.96)	19.01 (14.10–25.62)	0.17 (0.05–0.62)
61.3	2-y follow-up	0.86 (0.67–0.96)	0.94 (0.94–0.95)	15.46 (12.94–18.48)	0.15 (0.06–0.37)
45.2	2-y follow-up	0.87 (0.78–0.93)	0.95 (0.95–0.95)	17.00 (15.44–18.73)	0.14 (0.08–0.24)
53.4	Colonoscopy	0.53 (0.29–0.76)	0.87 (0.86–0.88)	4.10 (2.65–6.35)	0.54 (0.34–0.87)
59.3	Colonoscopy	0.77 (0.46–0.95)	0.94 (0.92–0.95)	12.13 (8.10–18.18)	0.25 (0.09–0.66)
60.0	Colonoscopy	0.75 (0.35–0.97)	0.95 (0.93–0.96)	14.40 (9.05–22.92)	0.26 (0.08–0.88)
62.7	2-y follow-up	1.00 (0.77–1.00)	0.93 (0.91–0.94)	13.01 (10.75–15.74)	0.04 (0.01–0.55)
58.8	Colonoscopy	0.85 (0.54–0.97)	0.92 (0.91–0.92)	10.21 (8.02–13.01)	0.17 (0.05–0.60)
49.0	Colonoscopy	0.96 (0.82–1.00)	0.87 (0.85–0.88)	7.21 (6.41–8.12)	0.04 (0.01–0.28)
62.7	Colonoscopy	0.73 (0.45–0.92)	0.96 (0.95–0.96)	16.44 (11.46–23.59)	0.28 (0.12–0.65)
62.7	Colonoscopy	0.60 (0.32–0.84)	0.95 (0.94–0.96)	13.06 (8.29–20.57)	0.42 (0.23–0.78)

CRC = colorectal cancer; FIT = fecal immunochemical test; LR = likelihood ratio.

*Either a colonoscopy (detects CRC and adenomas) or a 2-y longitudinal follow-up using a cancer registry (only detects CRC) was used for patients with negative FIT results.

†Discontinued and no longer in production in the United States.

‡Mean age >45 y because inclusion criteria for patients had to be ages >50 y.

§Mean age >45 y because only 21% of their cohort participants were aged 40–49 y.

Table 2

Subgroup Analysis Based on FIT Samples, FIT Cutoff Value, FIT Brand, and Reference Standard

Variable	Studies, *n*	Sensitivity (95% CI)	I^2*	Between-Study Variance in Logit Sensitivity	Specificity (95% CI)	I^2*	Between-Study Variance in Logit Specificity	Positive LR (95% CI)	I^2*	Negative LR (95% CI)	I^2*
FIT samples											
1-sample	13	0.78 (0.65–0.87)	75.9	1.00	0.95 (0.93–0.96)	98.9	0.33	14.2 (11.6–17.5)	91.6	0.24 (0.15–0.38)	84.8
2-sample	4	0.77 (0.59–0.89)	61.1	0.43	0.93 (0.90–0.95)	98.7	0.17	11.2 (6.5–19.5)	89.0	0.25 (0.13–0.49)	73.9
3-sample	6	0.80 (0.66–0.89)	5.2	0.07	0.93 (0.89–0.95)	97.7	0.30	11.3 (7.4–17.5)	88.4	0.21 (0.12–0.38)	0
FIT cutoff value for a positive test result											
<20 μg/g	11	0.86 (0.75–0.92)	56.4	0.60	0.91 (0.89–0.93)	98.6	0.16	9.8 (7.7–12.5)	93.3	0.16 (0.09–0.28)	62.2
20–50 μg/g	6	0.63 (0.43–0.79)	56.2	0.60	0.96 (0.94–0.97)	94.7	0.25	16.6 (12.9–21.4)	0	0.39 (0.24–0.63)	73.1
>50 μg/g	4	0.67 (0.59–0.74)	33.2	0.00	0.96 (0.94–0.98)	99.2	0.24	18.7 (11.7–29.8)	92.2	0.34 (0.27–0.43)	37.6
FIT brand											
OC-Micro/Sensor	5	0.86 (0.68–0.95)	0	0.28	0.91 (0.87–0.94)	95.5	0.21	9.7 (6.8–13.9)	54.4	0.16 (0.06–0.38)	0
OC-Light	4	0.93 (0.83–0.97)	26.6	0.07	0.91 (0.88–0.92)	95.9	0.06	9.9 (8.0–12.2)	85.7	0.08 (0.03–0.20)	9.99
Reference standard											
Colonoscopy	12	0.71 (0.58–0.81)	64.4	0.66	0.94 (0.91–0.96)	98.8	0.47	11.4 (8.6–15.2)	82.1	0.31 (0.21–0.45)	74.0
2-y follow-up†	7	0.87 (0.75–0.93)	39.0	0.39	0.94 (0.92–0.96)	97.1	0.18	15.2 (11.6–20.0)	85.4	0.14 (0.07–0.27)	29.5

FIT = fecal immunochemical test; LR = likelihood ratio.

*Inconsistency index minus the measure of heterogeneity.

†At least a 2-*y* longitudinal follow-up with medical records or cancer registry.

Table 3

Sensitivity Analysis: Summary Estimates of Subgroups After Removing Discontinued FITs

Variable	Studies, *n*	Sensitivity (95% CI)	I^2*	Between-Study Variance in Logit Sensitivity	Specificity (95% CI)	I^2*	Between-Study Variance in Logit Specificity	Positive LR (95% CI)	I^2*	Negative LR (95% CI)	I^2*
FIT sample											
1-sample	11	0.78 (0.65–0.87)	58.1	0.63	0.94 (0.92–0.95)	98.2	0.17	12.8 (10.8–15.1)	85.9	0.23 (0.14–0.38)	49.8
2-sample†	4	0.77 (0.59–0.89)	61.1	0.43	0.93 (0.90–0.95)	98.7	0.17	11.2 (6.5–19.5)	89.0	0.25 (0.13–0.49)	73.9
3-sample	6	0.80 (0.66–0.89)	5.2	0.07	0.93 (0.89–0.95)	97.7	0.30	11.3 (7.4–17.5)	88.4	0.21 (0.12–0.38)	0
FIT cutoff value for a positive test result											
<20 μg/g	9	0.89 (0.80–0.95)	26.4	0.32	0.91 (0.89–0.93)	94.9	0.12	10.2 (8.3–12.3)	75.2	0.12 (0.06–0.22)	14.5
20–50 μg/g	5	0.70 (0.55–0.81)	0	0.10	0.95 (0.95–0.96)	82.0	0.03	15.3 (12.5–18.8)	0	0.32 (0.21–0.49)	0
>50 μg/g	4	0.67 (0.59–0.74)	33.2	0.00	0.96 (0.94–0.98)	99.2	0.24	18.7 (11.7–29.8)	92.2	0.34 (0.27–0.43)	37.6
FIT brand											
OC-Micro/Sensor	5	0.86 (0.68–0.95)	0	0.28	0.91 (0.87–0.94)	95.5	0.21	9.7 (6.8–13.9)	54.4	0.16 (0.06–0.38)	0
OC-Light	4	0.93 (0.83–0.97)	26.6	0.07	0.91 (0.88–0.92)	95.9	0.06	9.9 (8.0–12.2)	85.7	0.08 (0.03–0.20)	9.99
Reference standard											
Colonoscopy	10	0.77 (0.65–0.86)	45.7	0.60	0.93 (0.91–0.95)	98.2	0.21	11.6 (9.6–14.0)	78.9	0.25 (0.16–0.39)	27.1
2-y follow-up‡	5	0.91 (0.78–0.97)	0	0.52	0.94 (0.91–0.96)	97.9	0.25	15.6 (10.8–22.7)	88.0	0.09 (0.03–0.25)	0

FIT = fecal immunochemical test; LR = likelihood ratio.

*Inconsistency index minus the measure of heterogeneity.

†Unable to do a sensitivity analysis because of the lack of data sets/studies.

‡At least a 2-y longitudinal follow-up with medical records or cancer registry.