Systematic review and bivariate/HSROC random-effect meta-analysis of immunochemical and guaiac-based fecal occult blood tests for colorectal cancer screening

Robert Launois[a], Jean-Gabriel Le Moine[a], Bernard Uzzan[b], Lucia I. Fiestas Navarrete[a] and Robert Benamouzig[b]

Background Current literature evidences higher accuracy of immunological (iFOBT) vis-à-vis guaiac-based (gFOBT) fecal occult blood tests for colorectal cancer (CRC) screening. Few well-designed head-to-head comparisons exist.

Aim This meta-analysis assesses the performances of two iFOBTs compared with an established gFOBT using colonoscopy as the gold standard.

Methods We mobilized a bivariate and a hierarchical summary receiver operating characteristic (HSROC) model. Positive likelihood ratio (LR^+) and negative likelihood ratio (LR^-) and diagnostic odds ratios were back-calculated. We constructed bivariate credibility ellipses in the HSROC space and calculated areas under the curve to obtain a global measure of test performance. Estimates are presented at 95% credibility levels.

Results We included and analyzed 21 studies. OC-Sensor was the best test for CRC screening, with high sensitivity (0.87; 95% credibility interval: 0.73–0.95) and specificity (0.93; 95% credibility interval: 0.84–0.96), optimal LR^+ (12.01) and LR^- (0.14), and a high diagnostic odds ratio (88.05). Bivariate credibility ellipses showed OC-Sensor's dominance over Hemoccult (sensitivity: 0.47; 95% credibility interval: 0.37–0.58; specificity: 0.93; 95% credibility interval: 0.91–0.95).

Conclusion Our findings support the use of OC-Sensor for CRC detection. The diagnostic estimates obtained may be extended to derive model parameters for economic decision models and to offer insight for future clinical and public health decision making. Our findings could influence the future of FOBTs within the CRC screening arsenal. *Eur J Gastroenterol Hepatol* 00:000–000 © 2014 Wolters Kluwer Health | Lippincott Williams & Wilkins.

European Journal of Gastroenterology & Hepatology 2014, **00**:000–000

Keywords: accuracy, advanced adenomas, bivariate, colorectal cancer, diagnostic test, HSROC, literature review, meta-analysis, QUADAS

[a]French Network for Evaluation in Health Economics, REES-France, Paris and [b]Service d'Hépato-Gastro-entérologie, Hôpital Avicenne APHP, Bobigny, France

Correspondence to Robert Launois, PhD, French Network for Evaluation in Health Economics, REES-France, 28 Rue d'Assas, 75006 Paris, France
Tel: + 33 1 44 39 16 90; e-mail: launois.rees@wanadoo.fr

Received 16 April 2014 **Accepted** 16 June 2014

Introduction

Each year, 320 000 new cancer cases are diagnosed in France [1]. With 40 000 new cases per year, colorectal cancer (CRC) has one of the highest incidences observed in the French population, ranking third after prostate (62 000) and breast cancer (50 000). It is estimated that at least 17 000 French people die from CRC each year; as such, it is the second largest cause of mortality among women and the third among men.

In 2005, 25 600 CRC patients benefited from a 100% reimbursement of healthcare costs associated with long-term care schemes [1]. This evidences the important economic repercussions that the management and treatment of the disease could pose on the French public insurance accounts [2]. An extensive body of research shows the effectiveness of CRC screening strategies on disease prevention [3–6], elucidating their life-saving and cost-saving potential. In fact, the advancement and increasing number of CRC screening techniques utilized in France point to a shift of priorities toward early detection.

Proposed biennially to ~17 million individuals aged 50–74 years old, nonrehydrated Hemoccult has been the established screening test of choice to detect CRC in an average-risk population since 2002 in France. Implementation of screening alternatives using immunological tests has been proposed to overcome the main limitations of the guaiac-based tests, namely, low sensitivity, qualitative reading, and low specificity for human hemoglobin. However, few well-designed head-to-head comparisons exist [7].

As such, the comparative effectiveness assessment of CRC screening technologies in current use in France is the best time for the debate. The present meta-analysis aims to assess the performances of two immunochemically based fecal occult blood tests (i.e. OC-Sensor and Magstream) compared with an established guaiac-based fecal occult blood test (i.e. Hemoccult) using colonoscopy as the gold standard. This meta-analysis did not consider other aspects contributing toward the choice of a screening test: adherence of individuals to testing and participation rate, sample strategy, and sample logistics.

DOI: 10.1097/MEG.0000000000000160

In this respect, our objective is only to bridge the gaps in the existing body of evidence in terms of the screening accuracy of immunochemical and guaiac tests for the detection of advanced adenoma and CRC.

Methods

We carried out this study in accordance with the standards set forth by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [8].

We used the PICOS criteria to guide the scope of the literature review and construct the search equation. The following five PICOS components provided a framework for our research question and facilitated the database search process: characteristics of the patient population (P), nature of the intervention (I), selected comparators (C), outcome measurements (O), and study design (S) [8]. We searched PubMed and EMBASE from 1980 to 2013 and the Cochrane Central Register of Controlled Trials from inception to the last quarter of 2012. Only English and French language articles were searched. The search was performed on 1 October 2013.

Articles were included in the meta-analysis if they fulfilled all of the following criteria: (i) study patients were 40 years of age or older (A number of trials were excluded because of the age of participants. Studies that included participants younger than 40 years of age were only included in the meta-analysis if the mean age of the included population was over 40 years. There was no superior limit for age.), with an average risk of CRC (i.e. no family history of cancer, no indication for CRC screening, and no indication of positive screening for CRC), and without having undergone any CRC screening over the last 6 months, (ii) the screening intervention included either nonrehydrated Hemoccult, Magstream, or OC-Sensor, (iii) the reference tests used were either colonoscopy for all cases, colonoscopy for positive tests and follow-up registry for negative tests, or colonoscopy for positive tests and sigmoidoscopy for negative tests (lower gastrointestinal tract endoscopy was deemed equivalent to a sigmoidoscopy), (iv) the findings presented enabled the calculation of true positives (TPs), true negatives (TNs), false positives (FPs), and false negatives (FNs), and (v) the study followed either a single-gate or a two-gate design.

The conditions of interest were advanced adenoma and (CRC. Advanced adenoma was defined as an adenoma with a size superior to 10 mm and/or the presence of a villous adenoma with a greater than 20% villous architecture, and/or the presence of high-grade dysplasia. All-stage CRCs were taken into account. Studies that performed partial verification were excluded, even if specificity estimates were calculated using the rare disease hypothesis [9].

The screening interventions of interest included one guaiac-based test, known as Hemoccult (Beckman Coulter Inc., Fullerton, California, USA), and two immunochemical-based tests consisting of Magstream (Fujirebio Inc., Japan) and OC-Sensor (Eiken Chemical Co. Ltd, Japan). As several versions of these tests are available, we opted to adopt the following conventions delimiting the test modalities that each of them includes. There are currently three versions of the Hemoccult test in use: Hemoccult, Hemoccult II, and Hemoccult Sensa. We included only the nonrehydrated modalities of the test (i.e. Hemoccult, Hemoccult II) and considered them together as 'Hemoccult' throughout the present study [10]. Hemoccult Sensa (The Hemoccult Sensa test was not considered for this meta-analysis as it is not used in France. The combination of the results of Hemoccult II and Hemoccult Sensa was not included either as it did not make sense to introduce more heterogeneity into the analysis, Sensa being more sensitive.) was not included in our meta-analyses. Moreover, the Immudia Hem/SP test is available in either one of two modalities: quantitative Magstream and semiquantitative HemeSelect. As such, we used the most recent appellation (i.e. 'Magstream') to refer to both of these test categories throughout our work [10,11]. Finally, for the OC test series, we included the most utilized versions of the test: OC-Light, OC-Hemodia, OC-Micro, and OC-Sensor. Guided by current conventions used by the AHRQ [11], the CRD [10], and the INESSS [12], we considered the aforementioned tests to be equivalent and used the common denomination 'OC-Sensor' to refer to them throughout this meta-analysis.

AQ1

Studies were classified as having a single-gate design when they included participants in whom the disease status was unknown and compared the diagnostic results obtained with the index test against those obtained with the reference test [10]. Such a single-gate design is typical of diagnostic and longitudinal cohort studies. The main difference lies in the time interval between the administration of the index test and the reference standard. Diagnostic cohort studies tend to administer both tests simultaneously or soon after one another. In contrast, longitudinal cohort studies perform the index test a priori and proceed to follow patients through time until the disease of interest becomes evident [13]. However, studies were classified as having a two-gate design when they estimated the sensitivity of the index test in patients who had an established diagnosis and the specificity of the same test among healthy controls [10]. Such a design is typical of diagnostic case–control studies. Although single-gate studies are preferred over two-gate designs, as they are likely to represent a realistic clinical practice scenario [13], we opted to extend our inclusion criteria to both single-gate and two-gate designs. Following this reasoning, diagnostic cohort studies, longitudinal cohort studies, and case–control studies could be included in the meta-analysis.

All qualifying studies were assessed on the basis of the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) protocol [14] using the Cochrane's computer program Review Manager (RevMan, version 5.2.6; The Nordic Cochrane Centre, The Cochrane Collaboration, 2012, Copenhagen, Denmark). QUADAS is an evidence-based quality assessment tool that is structured as a list of 14 items, 11 of which are incorporated into the RevMan analysis. Each item is formulated to be answered as either 'yes', 'no', or 'unclear', indicating a high, a low, and an unclear risk of bias (Table 1).

All data were extracted in duplicate by two investigators using a standard protocol and reporting form. Disagreements were resolved by a third investigator. For every study, the number of TPs, TNs, FPs, and FNs was retrieved and documented. Sensitivity and specificity were then calculated for CRC and advanced adenoma screening, when available. In addition, we collected information on the name of the study, year of publication, number of patients, type of screening technique used, nature of the comparator, as well as inclusion and exclusion criteria. We could extract overall study data, without the need to obtain individual-level data.

Positioned at the center of diagnostic theory, sensitivity and specificity are the preferred measures used in meta-analyses of screening accuracy, given that they estimate a test's ability to correctly classify individuals as diseased or disease-free. Moreover, they allow for the back-calculation of other summary estimates, including likelihood and diagnostic odds ratios (DORs). Specificity was defined taking into consideration only the lesions of interest (i.e. for detection of advanced adenomas, CRCs were considered FPs). We used two hierarchical logistic regression models: a bivariate model and a hierarchical summary receiver operating characteristic (HSROC) model, which respect the binomial structure of the data and account for between-study heterogeneity [15].

Table 1 **Items relevant to the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) protocol**

(1) Was the spectrum of patients' representative of the patients who will receive the test in practice? (representative spectrum)
(2) Is the reference standard likely to classify the target condition correctly? (acceptable reference standard)
(3) Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests? (acceptable delay between tests)
(4) Did the whole sample or a random selection of the sample, receive verification using the intended reference standard? (partial verification avoided)
(5) Did patients receive the same reference standard irrespective of the index test result? (differential verification avoided)
AQ2 (6) Was the reference standard independent of the index test
(7) Were the reference standard results interpreted without knowledge of the results of the index test? (index test results blinded)
(8) Were the index test results interpreted without knowledge of the results of the reference standard? (reference standard results blinded)
(9) Were the same clinical data available when test results were interpreted as would be available when the test is used in practice? (relevant clinical information)
(10) Were uninterpretable/intermediate test results reported? (uninterpretable results reported)
(11) Were withdrawals from the study explained? (withdrawals explained)

We chose to use the bivariate and HSROC models in view of the known limitations with the use of the Littenberg–Moses summary receiver operating characteristic (ROC) curve in meta-analyses of screening accuracy [15,16]. Moreover, because of its random-effect approach, the bivariate/HSROC method allows for the incorporation of variability into the analysis [17]. This was particularly important to the design of our study considering the differences in implicit thresholds that we would expect across the studies included. As stated by Sutton: 'If all or a proportion of heterogeneity is not explainable, then it needs to be allowed for in the analysis. This is commonly done in meta-analysis by incorporating random effects into the synthesis models' [18].

The bivariate model uses a random-effects approach in the estimation of summary points for sensitivity and specificity as well as in the estimation of 95% credibility intervals. The method is based on modeling (logit) sensitivity and specificity as bivariate normal distributions. The logit-transformed sensitivity in study *i* is assumed to have a mean of $\mu_{A,i}$, whereas the true logit sensitivities of individual studies ($\mu_{A,i}$) are assumed to be distributed around a common mean value μ_A and have a within-study variability of σ_A^2. However, the true logit specificities of individual studies ($\mu_{B,i}$) are assumed to have a mean value of μ_B and a between-study variance of σ_B^2. The correlation parameter is obtained from the two posterior means of the two univariate sensitivity and specificity models, which are obtained using empirical Bayes predictions. As such, the model includes five parameters (i.e. μ_A, σ_A^2, μ_B, σ_B^2, and σ_{AB}) leading to:

$$\begin{pmatrix} \mu_{Ai} \\ \mu_{Bi} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}, \sum_{AB} \right) \text{with} \sum_{AB} = \begin{pmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{pmatrix}.$$

We then calculated the subsequent measures of interest:

The positive and negative likelihood ratios, represented by LR^+ and LR^-, respectively,

$$\mathrm{LR}^+ = \frac{\mathrm{e}^{\mu_A}/(1+\mathrm{e}^{\mu_A})}{1-\{\mathrm{e}^{\mu B}/(1+\mathrm{e}^{\mu B})\}},$$

$$\mathrm{LR}^- = \frac{1-\{\mathrm{e}^{\mu A}/(1+\mathrm{e}^{\mu A})\}}{\mathrm{e}^{\mu_B}/(1+\mathrm{e}^{\mu_B})}.$$

The DOR defined by

$$\mathrm{DOR} = \mathrm{e}^{(\mu_A+\mu_B)}.$$

Upon fitting the bivariate model, we proceeded to transform the parameter estimates from the bivariate model into those of the HSROC model using the delta method [19].

The HSROC model [15] estimates the probability π_{ij} that a patient in a study i with disease status j has a positive test result, where $j=0$ for a patient without the disease and $j=1$ for a patient presenting the disease.

The HSROC model for study i is

$$\mathrm{logit}(\pi_{ij}) = (\theta_i+\alpha_i X_{ij})\exp(-\beta X_{ij}).$$

where α characterizes the accuracy parameter and θ is the positivity threshold parameter, which are assumed to vary between studies and have independent normal distributions. In addition, $X_{ij}=-1/2$ for disease-free individuals and $+1/2$ for those presenting the disease.

This model allowed the development of an HSROC curve by holding the accuracy parameter, α_i, fixed at its mean, Λ, while allowing the threshold parameter, θ_i, to vary. Thus, specificity could be estimated from a given sensitivity [15,20]:

$$\mathrm{logit}\,(\mathrm{sensitivity}) = \Lambda e^{-\beta/2} - e^{-\beta}\mathrm{logit}\,(\mathrm{specificity}).$$

Bivariate credibility regions were then constructed in the HSROC space. The ellipses denoting the joint credibility region for the means of logit-transformed sensitivity and specificity, μ_A and μ_B, were estimated per screening modality using the following formulas:

$$\mu_A = \hat{\mu}_A+\hat{s}_A\times c\times \cos(t),$$

$$\mu_B = \hat{\mu}_B+\hat{s}_B\times c\times \cos(t+\arccos(\hat{r})),$$

where $\hat{\mu}_A$ and $\hat{\mu}_B$ correspond to the posterior estimates of μ_A and μ_B, $\hat{s}_A$ and $\hat{s}_B$ are the associated standard errors, and $\hat{r}$ is an estimate of the correlation between $\hat{\mu}_A$ and $\hat{\mu}_B$. Finally, t takes values between 0 and $2\eth$, and c represents the boundary constant of the ellipse. c is defined by $c=\sqrt{\chi^2_{2,\alpha}}$, where $\chi^2_{2,\alpha}$ and is sampled from a χ^2 distribution with two degrees of freedom.

We calculated the area under the curve (AUC) by trapezoidal integration to obtain a global measure of test performance. We used the guidelines suggested by Swets [21] for the interpretation of intermediate AUC values, thus categorizing the observed values within the low ($0.5\geq \mathrm{AUC}\leq 0.7$), moderate ($0.7\geq \mathrm{AUC}\leq 0.9$), and high ($0.9\geq \mathrm{AUC}\leq 1$) screening accuracy ranges.

Results

The PRISMA flowchart is shown in Fig. 1. Our search identified 953 records: 761 of them were identified through database searches and an additional 192 through reports published by HTA bodies. Having removed all duplicates, our search identified 855 studies, of which 148 were relevant on the basis of their title and abstract and 22 fulfilled the predetermined selection criteria [7, 22–42]. Hence, we included 22 studies in the qualitative synthesis and meta-analysis. Figure 2 presents the quality assessment findings for the 22 studies included.

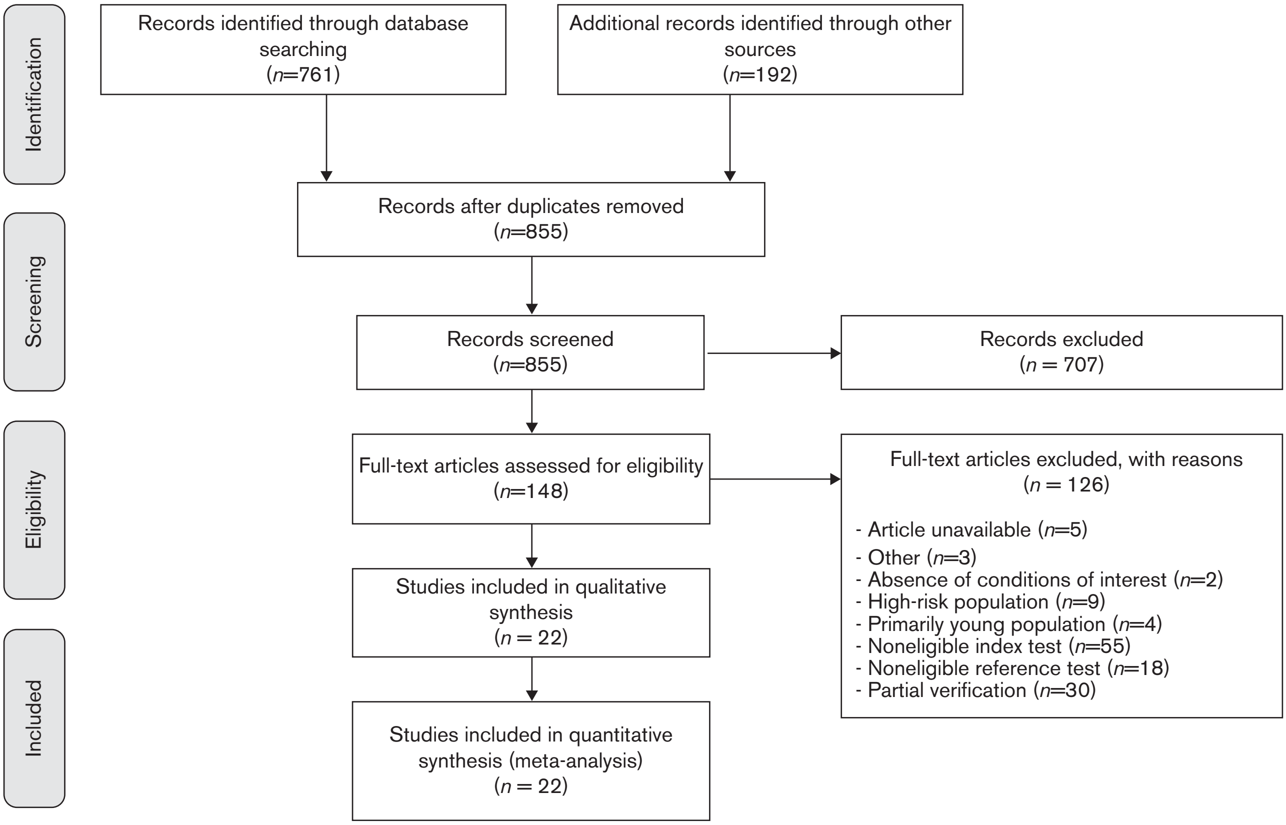
Among the 22 studies included, 17 were single-gate studies and five were two-gate studies. These included 11 diagnostic cohort studies, six longitudinal cohort studies, and five diagnostic case–control studies. These were published between the years 1992 and 2013. Twenty-two studies reported accuracy estimates for CRC screening, including eight studies that used Hemoccult, 10 that used OC-Sensor, and four that used Magstream. Fifteen studies reported accuracy estimates for advanced adenoma screening, including six studies that used Hemoccult, seven that used the OC-Sensor, and four that used Magstream. The total number of patients screened for advanced adenoma was 114 764 and the total number of patients screened for CRC was 174 469.

The screening accuracy analysis for advanced adenoma showed that 8–31% of patients screened with Hemoccult, 22–67% of patients screened with Magstream, and 15–62% of patients screened with the OC-Sensor obtained a TP diagnosis (Fig. 3). With respect to CRC detection, the analysis showed that 25–85% of patients screened with Hemoccult, 61–100% of patients screened with Magstream, and 26–100% of patients screened with OC-Sensor obtained a TP diagnosis (Fig. 4).

Through a graphical examination of the forest plots, we could determine that the results of the study by St John *et al.* [41] varied significantly from other studies using the same Magstream screening method. For this reason, the study by St John and colleagues was excluded from any further analysis.

Table 2 presents the summary estimates of sensitivity, specificity, LR$^+$, LR$^-$, and DOR obtained from the bivariate model for each screening modality and condition of interest. In the case of CRC, results show the OC-Sensor to have the best sensitivity among the three screening modalities analyzed. 87.2% of individuals presenting the disease are correctly identified as positive when using the OC-Sensor (sensitivity: 0.872; 95% credibility interval: 0.725–0.947) compared with 66.8% when using Magstream (sensitivity: 0.668; 95% credibility interval: 0.589–0.739) and 47.4% when using Hemoccult (sensitivity: 0.474; 95% credibility interval: 0.369–0.582). However, Magstream has the best specificity as 93.3% of individuals without the disease are correctly identified as negative when using the test (specificity: 0.933; 95% credibility interval: 0.917–0.945). It is worthwhile mentioning that both Hemoccult and the OC-Sensor have comparable specificities: 0.92 (95% credibility interval: 0.843–0.961) and 0.928 (95% credibility interval: 0.906–0.945), respectively.

Fig. 1


Identification
Screening
Eligibility
Included
Records identified through database searching (*n*=761)
Additional records identified through other sources (*n*=192)
Records after duplicates removed (*n*=855)
Records screened (*n*=855)
Records excluded (*n* = 707)
Full-text articles assessed for eligibility (*n*=148)
Full-text articles excluded, with reasons (*n* = 126)
- Article unavailable (*n*=5)
- Other (*n*=3)
- Absence of conditions of interest (*n*=2)
- High-risk population (*n*=9)
- Primarily young population (*n*=4)
- Noneligible index test (*n*=55)
- Noneligible reference test (*n*=18)
- Partial verification (*n*=30)
Studies included in qualitative synthesis (*n* = 22)
Studies included in quantitative synthesis (meta-analysis) (*n* = 22)

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram.

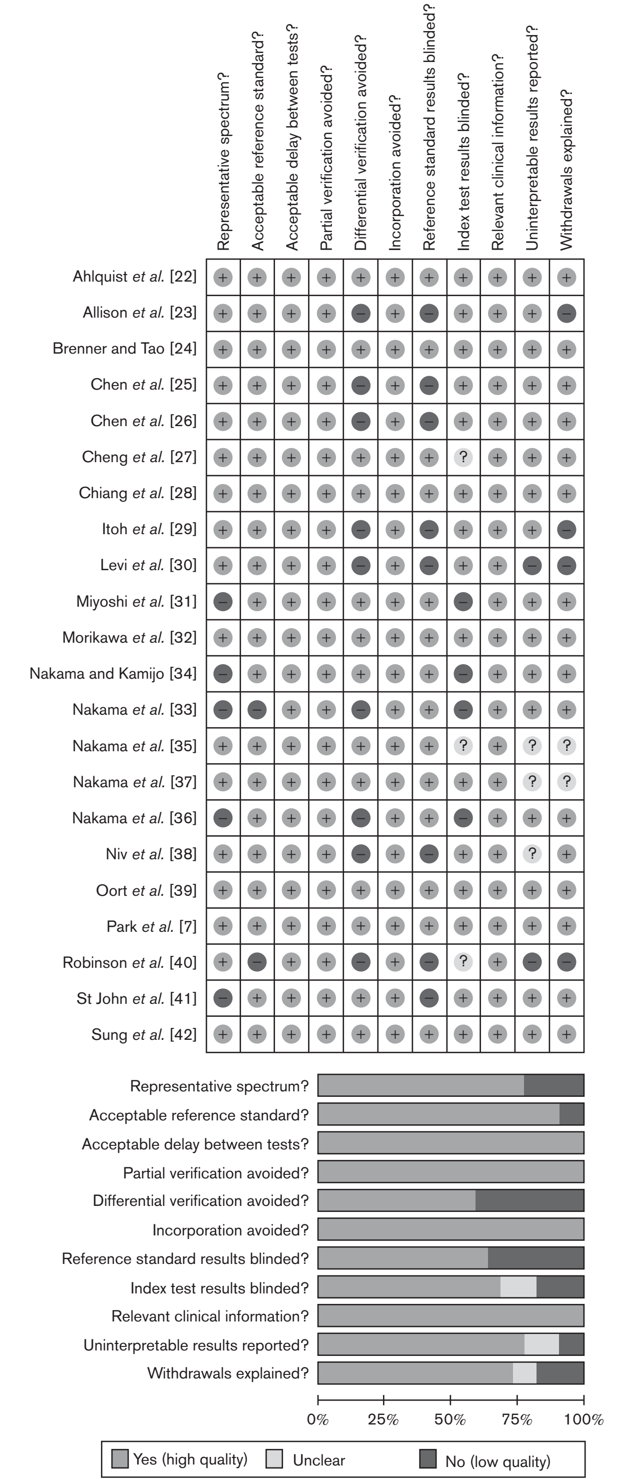
We found between-test differences in LR and DOR statistics. In this respect, the OC-Sensor is the best-performing test for CRC screening as it has the highest LR$^+$ (12.101) and the lowest LR$^-$ (0.137) among the three screening modalities. Thus, an individual who tests positive with OC-Sensor is 12 times more likely to have the disease than an individual with a negative test, whereas individuals who test negative with OC-Sensor are approximately seven times more likely (1/LR$^-$) to be disease-free than those with positive tests. Moreover, patients presenting with CRC are 88 times more likely to have a positive test with the OC-Sensor than disease-free individuals (DOR = 88.051).

Similar to screening for AdvAd, the summary estimates for sensitivity across screening modalities tend to be low, ranging from 0.142 to 0.477. However, the summary estimates for specificity are comparatively higher and range from 0.934 to 0.946. Magstream is the best-performing test, with the highest LR$^+$ (8.667) and the lowest LR$^-$ (0.553), although neither ratio is good enough to indicate that the test is informative. The DOR for Magstream shows that the positivity odds for patients with AdvAd are roughly 15 times greater than the positivity odds among patients without the condition.

Figures 5 and 6 show the pooled sensitivity and specificity estimates for the Hemoccult, Magstream, and OC-Sensor screening modalities for advanced adenoma and CRC, together with their corresponding 95% credibility ellipses represented in the ROC space. The ellipses indicate the area likely to contain the true mean test accuracy values of sensitivity and specificity for each screening modality. When screening for advanced adenoma, the ellipses do not show significant differences in sensitivity and specificity between Hemoccult, Magstream, and the OC-Sensor (Fig. 5). Conversely, similar to the differential accuracy of CRC screening modalities, Fig. 6 shows a clear difference between the sensitivity and the specificity of the OC-Sensor compared with Hemoccult: the OC-Sensor is significantly more accurate than Hemoccult. We did not find strong evidence for differences in accuracy between the OC-Sensor and Magstream or between Magstream and Hemoccult.

Following the significant results obtained by the bivariate ellipses, we constructed an HSROC plot for CRC screening modalities to better illustrate the expected diagnostic trade-off between sensitivity and specificity. We did not move forward with the HSROC analysis of advanced adenoma screening as the bivariate ellipses

Fig. 2



	Representative spectrum?	Acceptable reference standard?	Acceptable delay between tests?	Partial verification avoided?	Differential verification avoided?	Incorporation avoided?	Reference standard results blinded?	Index test results blinded?	Relevant clinical information?	Uninterpretable results reported?	Withdrawals explained?
Ahlquist *et al.* [22]	+	+	+	+	+	+	+	+	+	+	+
Allison *et al.* [23]	+	+	+	+	−	+	−	+	+	+	−
Brenner and Tao [24]	+	+	+	+	+	+	+	+	+	+	+
Chen *et al.* [25]	+	+	+	+	−	+	−	+	+	+	+
Chen *et al.* [26]	+	+	+	+	−	+	−	+	+	+	+
Cheng *et al.* [27]	+	+	+	+	+	+	+	?	+	+	+
Chiang *et al.* [28]	+	+	+	+	+	+	+	+	+	+	+
Itoh *et al.* [29]	+	+	+	+	−	+	−	+	+	+	−
Levi *et al.* [30]	+	+	+	+	−	+	−	+	+	−	−
Miyoshi *et al.* [31]	−	+	+	+	+	+	+	−	+	+	+
Morikawa *et al.* [32]	+	+	+	+	+	+	+	+	+	+	+
Nakama and Kamijo [34]	−	+	+	+	+	+	+	−	+	+	+
Nakama *et al.* [33]	−	−	+	+	−	+	+	−	+	+	+
Nakama *et al.* [35]	+	+	+	+	+	+	+	?	+	?	?
Nakama *et al.* [37]	+	+	+	+	+	+	+	+	+	?	?
Nakama *et al.* [36]	−	+	+	+	−	+	+	−	+	+	+
Niv *et al.* [38]	+	+	+	+	−	+	−	+	+	?	+
Oort *et al.* [39]	+	+	+	+	+	+	+	+	+	+	+
Park *et al.* [7]	+	+	+	+	+	+	+	+	+	+	+
Robinson *et al.* [40]	+	−	+	+	−	+	−	?	+	−	−
St John *et al.* [41]	−	+	+	+	+	+	−	+	+	+	+
Sung *et al.* [42]	+	+	+	+	+	+	+	+	+	+	+

Representative spectrum?
Acceptable reference standard?
Acceptable delay between tests?
Partial verification avoided?
Differential verification avoided?
Incorporation avoided?
Reference standard results blinded?
Index test results blinded?
Relevant clinical information?
Uninterpretable results reported?
Withdrawals explained?
0%
25%
50%
75%
100%
Yes (high quality)
Unclear
No (low quality)

Quality Assessment of Diagnostic Accuracy Studies (QUADAS) quality assessment of the 21 studies included.

showed no significant differences between the tests. Figure 7 shows the estimated HSROC curves and expected operating points for Hemoccult, Magstream, and the OC-Sensor with respect to CRC screening. We calculated the AUC for each screening modality and condition of interest. The AUC measures global screening accuracy by estimating the probability that a randomly chosen individual is correctly classified as diseased or disease-free. We found significant differences in the accuracy between the three CRC screening modalities. When used in CRC screening, the AUC analysis shows that OC-Sensor has a high accuracy (AUC = 0.95), Magstream has a moderate accuracy (AUC = 0.81), and Hemoccult has a low accuracy (AUC = 0.66). Our findings showed that a CRC patient who is screened with the OC-Sensor has a 95% probability to obtain a more abnormal test than a disease-free individual. Taking the credibility intervals of the AUC values into account, we concluded that the screening accuracy of the OC-Sensor is significantly higher than that of Magstream and Hemoccult. We found no evidence pointing to a statistically significant difference in screening accuracy between Magstream and Hemoccult.

Discussion

The aim of our study was to synthesize the accrued evidence on the accuracy of tests that are currently used for CRC screening in France.

The decision to phase-in immunological tests into the existing screening arsenal was mainly on the basis of the findings from six studies [3,6,7,32,43,44]. Four of them [3–6] reported the efficacy of OC-Sensor and Magstream in relation to reductions in CRC mortality. These were supplemented by two screening accuracy studies [7,32] showing the increased sensitivity of the OC-Sensor and Magstream vis-à-vis guaiac-based tests. However, it is important to note that there are a number of limitations to be considered when assessing the quality of the evidence reported by these works.

First, none of the studies confronted the diagnostic performance of the three screening products against each other. Second, four of the six studies used mortality estimates in the estimation of diagnostic efficacy, which are known to overestimate the benefits of screening techniques [45]. Third, half of the studies were diagnostic case–controls, which are prone to bias and are considered to produce inflated estimates of test accuracy [46,47]. As such, we deemed it plausible that this relatively small body of evidence could have overestimated the overall benefit and the sensitivities of OC-Sensor and Magstream in CRC detection.

Thus, the present work was initiated with the objective of bridging the gaps in the existing body of evidence for the screening accuracy of immunochemical and guaiac-based tests for the detection of advanced adenoma and CRC.

Our study found the OC-Sensor to be the best-performing test for CRC screening. This was evidenced

Fig. 3

Hemoccult

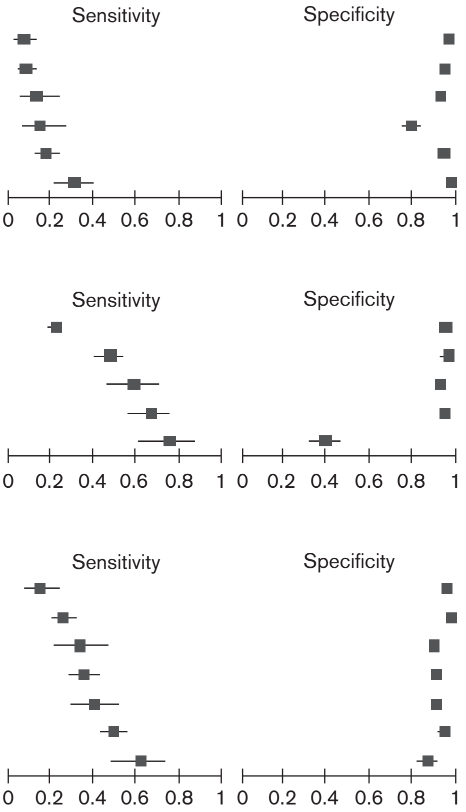
Study	TP	FP	FN	TN	Sensitivity	Specificity
Ahlquist *et al.* [22]	11	70	134	2282	0.08 (0.04–0.13)	0.97 (0.96–0.98)
Brenner and Tao [24]	19	92	203	1921	0.09 (0.05–0.13)	0.95 (0.94–0.96)
Park *et al.* [7]	8	53	51	648	0.14 (0.06–0.25)	0.92 (0.90–0.94)
Sung *et al.* [42]	9	92	50	354	0.15 (0.07–0.27)	0.79 (0.75–0.83)
Oort *et al.* [39]	35	87	159	1540	0.18 (0.13–0.24)	0.95 (0.93–0.96)
Allison *et al.* [23]	33	165	74	7793	0.31 (0.22–0.41)	0.98 (0.98–0.98)

Magstream

Study	TP	FP	FN	TN	Sensitivity	Specificity
Morikawa *et al.* [32]	145	1086	503	20071	0.22 (0.19–0.26)	0.95 (0.95–0.95)
Nakama *et al.* [33]	119	8	131	242	0.48 (0.41–0.54)	0.97 (0.94–0.99)
Nakama *et al.* [35]	41	745	29	9137	0.59 (0.46–0.70)	0.92 (0.92–0.93)
Allison *et al.* [23]	68	372	34	7019	0.67 (0.57–0.76)	0.95 (0.94–0.95)
St John *et al.* [41]	34	117	11	76	0.76 (0.60–0.87)	0.39 (0.32–0.47)

OC-Sensor

Study	TP	FP	FN	TN	Sensitivity	Specificity
Chen *et al.* [26]	12	1656	68	44256	0.15 (0.08–0.25)	0.96 (0.96–0.97)
Brenner and Tao [24]	57	53	165	1960	0.26 (0.20–0.32)	0.97 (0.97–0.98)
Park *et al.* [7]	20	67	39	644	0.34 (0.22–0.47)	0.91 (0.88–0.93)
Oort *et al.* [39]	69	145	125	1482	0.36 (0.29–0.43)	0.91 (0.90–0.92)
Cheng *et al.* [27]	31	652	46	6682	0.40 (0.29–0.52)	0.91 (0.90–0.92)
Nakama *et al.* [33]	123	11	127	239	0.49 (0.43–0.56)	0.96 (0.92–0.98)
Nakama *et al.* [36]	37	45	23	297	0.62 (0.48–0.74)	0.87 (0.83–0.90)



Forest plots presenting the punctual estimates of sensitivity and specificity and 95% credibility intervals of each study across three diagnostic tests for advanced adenoma. FN, false negative; FP, false positive; TN, true negative; TP, true positive.

by its sensitivity and specificity estimates (sensitivity: 0.87; specificity: 0.93) optimal positive and negative likelihood ratios ($LR^{+}=12.01$; $LR^{-}=0.14$) as well as a high DOR (88.05). Credibility regions for the summary sensitivity and specificity, obtained through bivariate analysis, showed the clear dominance of the OC-Sensor with respect to Hemoccult (sensitivity: 0.47; specificity: 0.92). We further confirmed significant differences in accuracy between CRC screening modalities through an AUC analysis. The OC-Sensor showed the highest screening accuracy (AUC = 0.95), followed by Magstream (AUC = 0.81) and Hemoccult (AUC = 0.66).

For advanced adenoma, the bivariate summary estimates for sensitivity were very low across the three screening modalities. Magstream, the test with the highest sensitivity for advanced adenoma detection, could only identify up to 37% of TPs. The generally low TP rates led to suboptimal values of likelihood ratios. Consequently, no test fell within the range that could identify it as sufficiently informative, leading to comparatively lower DORs across screening modalities. Moreover, the credibility ellipses obtained through the bivariate model could not show any clear differences in test accuracy between the three modalities when screening for advanced adenoma.

To our knowledge, this is the first meta-analysis that compares the diagnostic value of OC-Sensor, Magstream, and Hemoccult for the detection of advanced adenoma and CRC in an average-risk population. Nonetheless, we compared our findings with those of a recent meta-analysis by Whyte *et al.* [48] that synthesized the screening accuracy of the OC-Sensor for the detection of CRC. The specificity estimates obtained in our analysis (0.93) are in agreement with those obtained by Whyte and colleagues (0.97). However, the sensitivity estimates presented in this work (0.87) and those obtained by Whyte and colleagues (0.66) were moderately divergent. This could be attributed to important differences in the inclusion criteria between the two meta-analyses, namely, the fact that Whyte and colleagues included studies where the reference standard for positive results was not consistently colonoscopy.

Fig. 4

Hemoccult

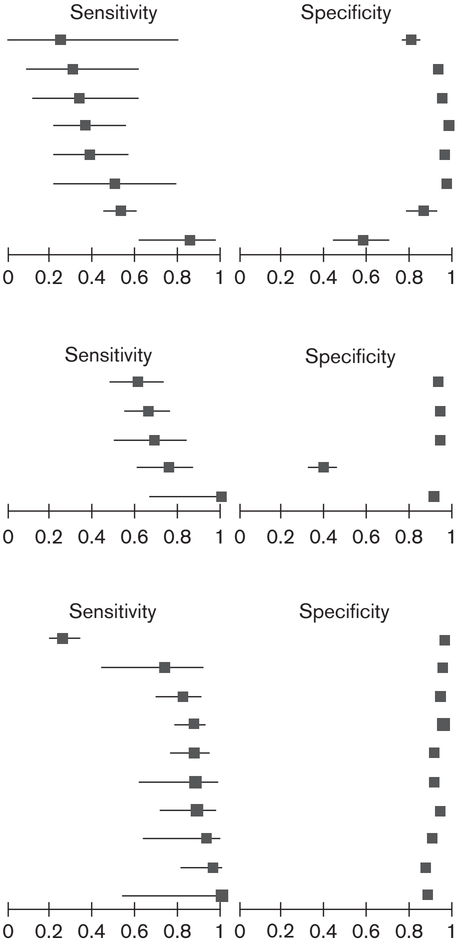
Study	TP	FP	FN	TN	Sensitivity	Specificity
Sung *et al.* [42]	1	100	3	401	0.25 (0.01–0.81)	0.80 (0.76–0.83)
Park *et al.* [7]	4	57	9	690	0.31 (0.09–0.61)	0.92 (0.90–0.94)
Brenner and Tao [24]	5	106	10	2114	0.33 (0.12–0.62)	0.95 (0.94–0.96)
Allison *et al.* [23]	13	185	22	7845	0.37 (0.21–0.55)	0.98 (0.97–0.98)
Niv *et al.* [38]	13	89	21	2145	0.38 (0.22–0.56)	0.96 (0.95–0.97)
Ahliquist *et al.* [22]	6	75	6	2410	0.50 (0.21–0.79)	0.97 (0.96–0.98)
Nakama and Kamijo [34]	106	14	94	86	0.53 (0.46–0.60)	0.86 (0.78–0.92)
Miyoshi *et al.* [31]	17	25	3	34	0.85 (0.62–0.97)	0.58 (0.44–0.70)

Magstream

Study	TP	FP	FN	TN	Sensitivity	Specificity
Nakama *et al.* [35]	39	747	25	9141	0.61 (0.48–0.73)	0.92 (0.92–0.93)
Morikawa *et al.* [32]	52	1179	27	20547	0.66 (0.54–0.76)	0.95 (0.94–0.95)
Allison *et al.* [23]	22	418	10	7043	0.69 (0.50–0.84)	0.94 (0.94–0.95)
St John *et al.* [41]	34	117	11	76	0.76 (0.60–0.87)	0.39 (0.32–0.47)
Robinson *et al.* [40]	9	136	0	1344	1.00 (0.66–1.00)	0.91 (0.89–0.92)

OC-Sensor

Study	TP	FP	FN	TN	Sensitivity	Specificity
Chen *et al.* [26]	39	1629	111	44213	0.26 (0.19–0.34)	0.96 (0.96–0.97)
Brenner ana Tao [24]	11	99	4	2121	0.73 (0.45–0.92)	0.96 (0.95–0.96)
Chen *et al.* [25]	53	1289	12	21318	0.82 (0.70–0.90)	0.94 (0.94–0.95)
Itoh *et al.* [29]	77	1413	12	26358	0.87 (0.78–0.93)	0.95 (0.95–0.95)
Oort *et al.* [39]	54	160	8	1599	0.87 (0.76–0.94)	0.91 (0.89–0.92)
Cheng *et al.* [27]	14	669	2	6726	0.88 (0.62–0.98)	0.91 (0.90–0.92)
Nakama *et al.* [35]	24	254	3	3979	0.89 (0.71–0.98)	0.94 (0.93–0.95)
Park *et al.* [7]	12	75	1	682	0.92 (0.64–1.00)	0.90 (0.88–0.92)
Chiang *et al.* [28]	27	370	1	2398	0.96 (0.82–1.00)	0.87 (0.85–0.88)
Levi *et al.* [30]	6	147	0	1051	1.00 (0.54–1.00)	0.88 (0.86–0.90)



Forest plots presenting the punctual estimates of sensitivity and specificity and 95% credibility intervals of each study across three diagnostic tests for colorectal cancer. FN, false negative; FP, false positive; TN, true negative; TP, true positive.

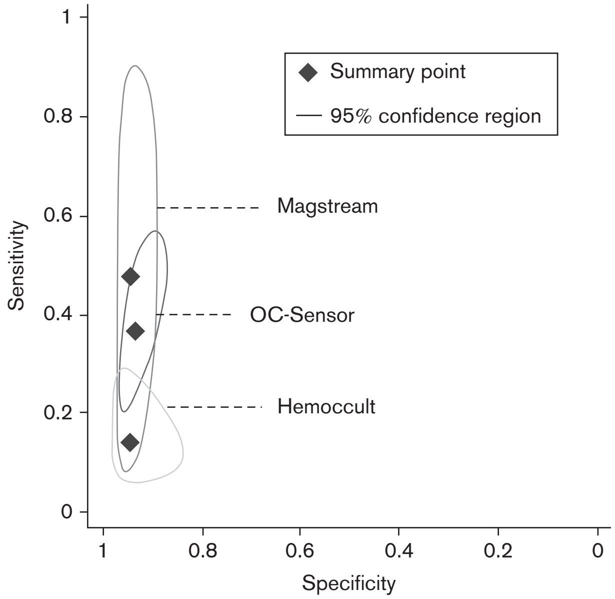
Table 2 Bivariate summary estimates of sensitivity, specificity, positive likelihood ratio, negative likelihood ratio, and diagnostic odds ratio for each screening modality and disease condition of interest

	Se	95% CI	Sp	95% CI	LR$^+$	LR$^-$	DOR
Screening modalities for advanced adenoma							
Hemoccult	0.142	0.092–0.211	0.946	0.902–0.971	2.612	0.908	2.878
Magstream	0.477	0.305–0.655	0.945	0.931–0.956	8.667	0.553	15.665
OC-Sensor	0.367	0.266–0.481	0.934	0.902–0.956	5.561	0.678	8.205
Screening modalities for colorectal cancer							
Hemoccult	0.474	0.369–0.582	0.92	0.843–0.961	5.944	0.571	10.400
Magstream	0.668	0.589–0.739	0.933	0.917–0.945	9.929	0.357	27.917
OC-Sensor	0.872	0.725–0.947	0.928	0.906–0.945	12.101	0.137	88.051

95% CI, credibility interval at 95%; DOR, diagnostic odds ratio; LR$^-$, negative likelihood ratio; LR$^+$, positive likelihood ratio; Se, sensitivity; Sp, specificity.

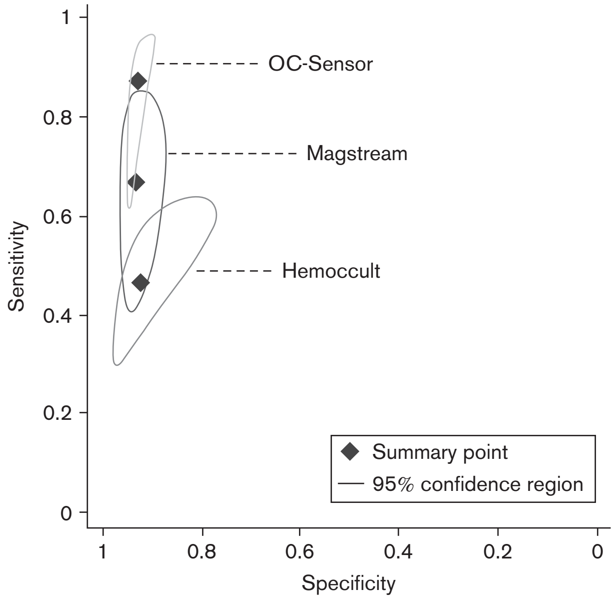
Our findings support the progressive phase-in of OC-Sensor tests in the French territory for CRC screening detection. We found no evidence to suggest that Magstream has significantly higher screening accuracy compared with Hemoccult. In this respect, our results do not support the National Cancer Screening Program's decision to progressively phase-in Magstream at the expense of Hemoccult's phase-out. Our comparative screening accuracy analysis indicates that only one (i.e. OC-Sensor) of the currently favored immunochemical screening alternatives overcomes the main limitations of the guaiac-based Hemoccult test in CRC detection.

Fig. 5



Bivariate summary estimates of sensitivity and specificity for each of the three screening modalities for advanced adenoma screening and the corresponding 95% credibility ellipse around the mean values. See Fig. 3 for primary data.

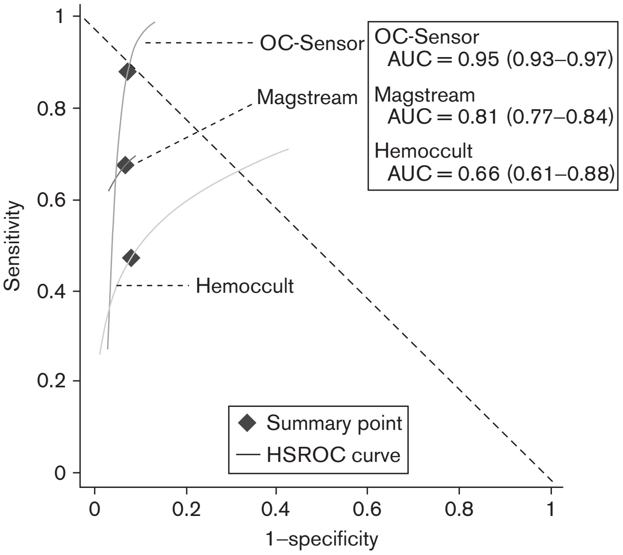
Fig. 6



Bivariate summary estimates of sensitivity and specificity for each of the three screening modalities for colorectal cancer screening and the corresponding 95% credibility ellipse around the mean values. See Fig. 4 for primary data.

Neither one of the three screening modalities analyzed proved to be significantly more accurate in the detection

Fig. 7



Estimated summary receiver operating characteristic curves and expected operating points for Hemoccult, Magstream, and the OC-Sensor for colorectal cancer screening on the basis of hierarchical regression modeling. Area under the curve (AUC) values and 95% credibility intervals are provided. HSROC, hierarchical summary receiver operating characteristic.

of advanced adenoma. As such, our results do not show the added benefit of using the OC-Sensor or Magstream, vis-à-vis Hemoccult, on early detection.

The assessment of screening accuracy is an important endeavor in and of itself. Yet, it should also be considered the foundational step from which to perform full economic evaluations by taking into account factors such as costs, side effects of tests, and consequences of correct classification and misclassification [18]. In this respect, the sensitivity and specificity estimates that we obtained in this meta-analysis may be extended to derive model parameters for health economic decision models for CRC screening. Of equal importance, our findings offer clinical insight for future screening practice. The back-calculated accuracy estimates produced throughout this work are of great practical use for clinical decision making, namely, the LR^+ and LR^-, DORs, and AUCs provided for each screening modality.

Our study has several strengths. First, the bivariate/HSROC approach guiding our analysis is the most statistically rigorous method in the diagnostic meta-analysis field. This is because of its ability to model the within-study binomial structure of the data while accounting for between-study heterogeneity [15,17]. Following a comparison of methods for meta-analysis of screening accuracy in current use, Centre for Reviews and Dissemination [16] conclude that the bivariate/HSROC method must be used as the standard, together with an analysis of summary ROC curves, credibility, and prediction regions. Indeed, ours is the first meta-analysis of screening accuracy ever carried out in France that uses this method.

Second, 17 of the 21 studies that we included in our meta-analysis corresponded to single-gate designs, which constitute the highest quality of evidence for diagnostic meta-analyses. This is because they better represent the clinical scenario where they would normally be used [10]. Third, we used a comprehensive search strategy to locate all relevant studies fulfilling our selection criteria. As such, we supplemented the search by identifying relevant article references from reports published by international HTA bodies. Moreover, four reviewers screened the retrieved studies in an effort to reduce the effect of publication bias. Finally, our meta-analysis was carried out using rigorous verification methods, particularly throughout the data extraction and quality assessment stages. On the one hand, data were extracted in duplicate by two investigators using a standard protocol and reporting form. On the other, the quality of each article was reviewed by two blinded raters using standard evaluation score sheets.

Findings must be interpreted in light of the following limitations. First, four of the 21 studies included in our meta-analysis were diagnostic case–controls [31,34,36, 49]. Their two-gate study design poses inherent problems in relation to spectrum bias. It is generally accepted that the selective inclusion of cases with more advanced disease tends to overestimate sensitivity and that the inclusion of healthy controls would lead to an overestimation of specificity. As such, it may well be argued that the four diagnostic case–controls included in the present work could have introduced an overestimation of the sensitivity and specificity of the tests. Because of the lack of available data, sensitivity analyses that exclude some of the references were not possible (i.e. decrease in precision and loss of significance), neither could a meta-regression be considered [The *Cochrane's handbook* explains that for such analyses a minimum of 10 studies is required (version 5.1.0, chapter 9.6.4).] [50]. Having acknowledged these limitations, it is relevant to add that these studies were used proportionately to estimate the sensitivity and specificity estimates across the three screening modalities.

Second, we included four studies whereby the inclusion criterion for patients' age was more than 20 years. We opted to include them having carefully assessed that the costs of excluding these studies, which were otherwise methodologically viable, were far too great. We made this concession considering that the mean age across the four studies ranged between 46.8 and 59.6 years. In these studies, the under 40 population varied between 2.5 and 27.24% of the total study population. Thus, we acknowledge that the population represented in our meta-analyses may not correspond completely to the standard 50–74-year-old population that is screened routinely. However, we defend our choice as it was necessary to reassess our inclusion criteria of patients' age to obtain a sufficient number of studies for synthesis.

Third, we included nine studies in which patients with a positive index test were verified using one reference standard and patients with a negative index were verified using a different standard. We acknowledge that this may have led to differential verification bias [51]. This bias could have been most important among the studies that used a colonoscopy for positive tests and follow-up registries for negative tests as the accuracy of the two methods is considerably different. If the patients testing positive receive a more accurate verification test than those testing negative, it is not improbable that an overestimation of the DORs may have occurred [19]. It is worth noting that differential bias could have impacted a rather small number of the studies included in our meta-analysis and that these studies were used proportionately across the screening modalities of interest.

For Magstream and the OC-Sensor, we chose to consider the recommended thresholds. However, for some studies, the threshold used differed from the one recommended. In this case, we included the closest threshold and used a random effect to address the resulting heterogeneity.

Finally, it is possible that our study could not find statistically significant differences between the OC-Sensor, Magstream, and Hemoccult in advanced adenoma detection because of the relatively low number of studies per subgroup meta-analysis. The number of studies varied between four and six; thus, analyses may have been underpowered to detect real differences in advanced adenoma screening accuracy. Although we found an important difference in screening accuracy between the OC-Sensor and Hemoccult, our findings did not find any significant differences between Magstream and either the OC-Sensor or Hemoccult. This could be because of the comparatively small number of studies that were used to estimate the sensitivity and specificity of Magstream. Thus, it cannot be rejected that the inclusion of a greater number of studies in the Magstream subgroup meta-analysis might have yielded statistically significant differences. This underlines the need for new screening data to narrow credibility intervals. The aforementioned are considered to be minor limitations.

Obviously, this meta-analysis did not consider explicitly the relevant advantages of immunochemical tests: the need for only one stool sample and absence of dietary or medication restrictions. These aspects could increase ease of use and participation. These advantages as well as the possibility for automation and customization of positivity according to colonoscopy have not been analyzed.

Conclusion

Our findings support the use of the OC-Sensor for CRC detection. The bivariate ellipse analysis showed the clear dominance of the OC-Sensor vis-à-vis Hemoccult, whereas the AUC analysis showed its high global test performance. We did not find significant differences in accuracy between Magstream and the OC-Sensor nor between Magstream and Hemoccult, pointing to the need for new diagnostic data to narrow credibility intervals. Our work bridges the gaps in the existing body of evidence on the accuracy of screening tests used currently for the detection of CRC and advanced adenoma in an average-risk population. The diagnostic estimates obtained here may be extended to derive model parameters for economic decision making as well as to offer insight for future clinical practice. As such, our findings have the potential to influence the near and longstanding future of fecal immunochemical test and guaiac-based fecal occult blood tests as part of the CRC screening arsenal.

Acknowledgements

Conflicts of interest

There are no conflicts of interest.

References

1 Belot A, Grosclaude P, Bossard N, Jougla E, Benhamou E, Delafosse P, *et al.* Cancer incidence and mortality in France over the period 1980–2005. *Rev Epidemiol Sante Publique* 2008; **56**:159–175.

2 Jeong KE, Cairns JA. Review of economic evidence in the prevention and early detection of colorectal cancer. *Health Econ Rev* 2013; **3**:20.

3 Nakajima M, Saito H, Soma Y, Sobue T, Tanaka M, Munakata A. Prevention of advanced colorectal cancer by screening using the immunochemical faecal occult blood test: a case–control study. *Br J Cancer* 2003; **89**:23–28.

4 Lee KJ, Inoue M, Otani T, Iwasaki M, Sasazuki S, Tsugane S. Japan Public Health Center-based Prospective Study. Colorectal cancer screening using fecal occult blood test and subsequent risk of colorectal cancer: a prospective cohort study in Japan. *Cancer Detect Prev* 2007; **31**:3–11.

5 Thosani N, Guha S, Singh H. Colonoscopy and colorectal cancer incidence and mortality. *Gastroenterol Clin North Am* 2013; **42**:619–637.

6 Nishihara R, Wu K, Lochhead P, Morikawa T, Liao X, Qian ZR, *et al.* Long-term colorectal-cancer incidence and mortality after lower endoscopy. *N Engl J Med* 2013; **369**:1095–1105.

7 Park DI, Ryu S, Kim YH, Lee SH, Lee CK, Eun CS, Han DS. 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**:2017–2025.

8 Moher D, Liberati A, Tetzlaff J, Altman DG. PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009; **339**:b2535.

9 Brecht JG, Robra BP. A graphic method of estimating the specificity of screening programmes from incomplete follow-up data. *Methods Inf Med* 1987; **26**:53–58.

10 Soares KV, Burch JA, Duffy S St, John DJ, Smith S, Westwood M, *et al. Diagnostic accuracy and cost-effectiveness of faecal occult blood tests used in screening for colorectal cancer: a systematic review*. London: Center for Research Dissemination; 2007.

11 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**:638–658.

12 Potvin E, Gosselin C. *Test immunochimique de recherche de sang occulte dans les selles. Détermination d'un seuil de positivité pour démarrer les projets de démonstration du PQDCCR*. Québec: Institut national d'excellence en santé et en services sociaux (INESSS); 2012.

13 Perez P, Saves M, Picat M-Q, Chene G. Methods in Clinical Research University Diploma: Massive Online Open Course; 2013.

14 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**:529–536.

15 Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med* 2001; **20**:2865–2884.

16 Centre for Reviews and Dissemination. *Systematic reviews: CRD's guidance for undertaking reviews in health care*. York: University of York; 2009.

17 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**:982–990.

18 Sutton AJ, Cooper NJ, Goodacre S, Stevenson M. Integration of meta-analysis and economic decision modeling for evaluating diagnostic tests. *Med Decis Making* 2008; **28**:650–667.

19 Cox C. Delta method. In: Armitage P, Colton T, editors. *Encyclopedia of biostatistics*. New York: Wiley; 1998. pp. 1125–1127.

20 Macaskill P. Empirical Bayes estimates generated in a hierarchical summary ROC analysis agreed closely with those of a full Bayesian analysis. *J Clin Epidemiol* 2004; **57**:925–932.

21 Swets JA. Measuring the accuracy of diagnostic systems. *Science* 1988; **240**:1285–1293.

22 Ahlquist DA, Sargent DJ, Loprinzi CL, Levin TR, Rex DK, Ahnen DJ, *et al.* Stool DNA and occult blood testing for screen detection of colorectal neoplasia. *Ann Intern Med* 2008; **149**:441–450.

23 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**:155–159.

24 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**:3049–3054.

25 Chen LS, Liao CS, Chang SH, Lai HC, Chen TH. Cost-effectiveness analysis for determining optimal cut-off of immunochemical faecal occult blood test for population-based colorectal cancer screening (KCIS 16). *J Med Screen* 2007; **14**:191–199.

26 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**:551–558.

27 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**:685–690.

28 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**:1474–1481.

29 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.

30 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; 2415–2424.

31 Miyoshi H, Ohshiba S, Asada S, Hirata I, Uchida K. Immunological determination of fecal hemoglobin and transferrin levels: a comparison with other fecal occult blood tests. *Am J Gastroenterol* 1992; **87**:67–73.

32 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**:422–428.

33 Nakama H, Fattah A, Zhang B, Uehara Y, Wang C. A comparative study of immunochemical fecal tests for detection of colorectal adenomatous polyps. *Hepatogastroenterology* 2000; **47**:386–389.

34 Nakama H, Kamijo N. Accuracy of immunological fecal occult blood testing for colorectal cancer screening. *Prev Med* 1994; **23**:309–313.

35 Nakama H, Zhang B, Fattah AA, Kamijo N, Zhang X. Characteristics of colorectal cancer that produce positive immunochemical occult blood test results on stool obtained by digital rectal examination. *Can J Gastroenterol* 2001; **15**:227–230.

36 Nakama H, Zhang B, Kamijo N. Sensitivity of immunochemical fecal occult blood test for colorectal flat adenomas. *Hepatogastroenterology* 2004; **51**:1333–1336.

37 Nakama H, Zhang B, Zhang X. Evaluation of the optimum cut-off point in immunochemical occult blood testing in screening for colorectal cancer. *Eur J Cancer* 2001; **37**:398–401.

38 Niv Y, Lev-El M, Fraser G, Abuksis G, Tamir A. Protective effect of faecal occult blood test screening for colorectal cancer: worse prognosis for screening refusers. *Gut* 2002; **50**:33–37.

39 Oort FA, Terhaar Sive Droste JS, Van Der Hulst RW, Van Heukelem HA, Loffeld RJ, Wesdorp IC, *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**:432–439.

40 Robinson MH, Marks CG, Farrands PA, Thomas WM, Hardcastle JD. Population screening for colorectal cancer: comparison between guaiac and immunological faecal occult blood tests. *Br J Surg* 1994; **81**:448–451.

41 St John DJ, Young GP, Alexeyeff MA, Deacon MC, Cuthbertson AM, Macrae FA, Penfold JC. Evaluation of new occult blood tests for detection of colorectal neoplasia. *Gastroenterology* 1993; **104**:1661–1668.

42 Sung JJ, Chan FK, Leung WK, Wu JC, Lau JY, Ching J, *et al*. Screening for colorectal cancer in Chinese: comparison of fecal occult blood test, flexible sigmoidoscopy, and colonoscopy. *Gastroenterology* 2003; **124**:608–614.

43 Saito H, Soma Y, Koeda J, Wada T, Kawaguchi H, Sobue T, *et al*. Reduction in risk of mortality from colorectal cancer by fecal occult blood screening with immunochemical hemagglutination test. A case–control study. *Int J Cancer* 1995; **61**:465–469.

44 Saito H, Soma Y, Nakajima M, Koeda J, Kawaguchi H, Kakizaki R, *et al*. A case–control study evaluating occult blood screening for colorectal cancer with hemoccult test and an immunochemical hemagglutination test. *Oncol Rep* 2000; **7**:815–819.

45 Launoy GD, Bertrand HJ, Berchi C, Talbourdet VY, Guizard AV, Bouvier VM, Caces ER. 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**:493–496.

46 Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH, Bossuyt PM. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999; **282**:1061–1066.

47 Westwood ME, Whiting PF, Kleijnen J. How does study quality affect the results of a diagnostic meta-analysis? *BMC Med Res Methodol* 2005; **5**:20.

48 Whyte S, Chilcott J, Cooper C, Essat M, Stevens J, Wong R, *et al*. *Re-appraisal of the options for colorectal cancer screening. Report for the NHS Bowel Cancer Screening Programme*. Sheffield: Scool of Health and related research (SCHARR); 2011.

49 Nakama H, Zhang B, Abdul Fattah AS, Kamijo N, Fukazawa K. Relationships between a sign of rectal bleeding and the results of an immunochemical occult blood test, and colorectal cancer. *Eur J Cancer Prev* 2000; **9**:325–328.

50 Deeks J, Higgins J, Altman DG. Analysing data and under taking meta-analyses. In: Higgins J, Green S, editors. *Cochrane handbook for systematic reviews of interventions, Cochrane Book Series*. Chichester, UK: John Wiley & Sons Ltd; 2008.

51 Schmidt RL, Factor RE. Understanding sources of bias in diagnostic accuracy studies. *Arch Pathol Lab Med* 2013; **137**:558–565.

AUTHOR QUERY FORM

LIPPINCOTT
WILLIAMS AND WILKINS

JOURNAL NAME: MEG

ARTICLE NO: ejgh8476

QUERIES AND / OR REMARKS

QUERY NO.	Details Required	Author's Response
Q1	Please provide city/town name for ‘Fujirebio Inc., Japan’ and ‘Eiken Chemical Co. Ltd, Japan’.	
Q2	Please confirm whether deletion of ‘(i.e.’ in item 6 is OK.	