A systematic review of the literature

June 2009

Comparison of diagnostic accuracy between immunochemical and guaiac based faecal occult blood tests for colorectal cancer detection

Arindam Basu
Pamela Smartt
This report should be referenced as follows:


Health Services Assessment Collaboration (HSAC), University of Canterbury

ISBN 978-0-9864544-3-1 (Online)
ISBN 978-0-9864544-4-8 (Print)
ISSN 1178-5748 (Online)
ISSN 1178-573X (Print)
Review Team

This review was undertaken by the Health Services Assessment Collaboration (HSAC). HSAC is a collaboration of the Health Sciences Centre of the University of Canterbury, New Zealand and Health Technology Analysts, Sydney, Australia. This report was authored by Dr Arindam Basu (Senior Researcher) and Dr Pamela Smartt (Senior Researcher) who developed and undertook the literature search, extracted the data, conducted the critical appraisals, and prepared the report.

Acknowledgements

Dr Ray Kirk (HSAC Director) and Professor James Allison (Clinical Professor of Medicine Emeritus, University of California at San Francisco, Division of Gastroenterology, San Francisco General Hospital) peer reviewed the final draft. Mr Stephen Lungley (Senior Policy Analyst, Cancer, Sector Capability and Innovation Directorate, New Zealand Ministry of Health) provided input to the draft at various stages. Ms Cecilia Tolan (HSAC Administrator) provided document formatting. Sub-editing was performed by Ms Jane Mountier.

Staff at the University of Canterbury Libraries assisted with the retrieval of articles.

The current review was conducted under the auspices of a contract funded by the New Zealand Ministry of Health. This report was requested by Mr Stephen Lungley, and Dr Brendon Grey (Clinical Director Acting, Bowel Cancer Screening, Cancer, Sector Capability and Innovation Directorate, New Zealand Ministry of Health). The systematic review of the evidence will ultimately be used in conjunction with other information to inform policy decision making. The content of the review alone does not constitute clinical advice or policy recommendations.

Copyright Statement & Disclaimer

This report is copyright. Apart from any use as permitted under the Copyright Act 1994, no part may be reproduced by any process without written permission from HSAC. Requests and inquiries concerning reproduction and rights should be directed to the Director, Health Services Assessment Collaboration, Health Sciences Centre, University of Canterbury, Private Bag 4800, Christchurch, New Zealand.

HSAC takes great care to ensure the accuracy of the information in this report, but neither HSAC, the University of Canterbury, Health Technology Analysts Pty Ltd nor the Ministry of Health make any representations or warranties in respect of the accuracy or quality of the information, or accept responsibility for the accuracy, correctness, completeness or use of this report.

The reader should always consult the original database from which each abstract is derived along with the original articles before making decisions based on a document or abstract. All responsibility for action based on any information in this report rests with the reader.
This report is not intended to be used as personal health advice. People seeking individual medical advice should contact their physician or health professional.

The views expressed in this report are those of HSAC and do not necessarily represent those of the University of Canterbury New Zealand, Health Technology Analysts Pty Ltd, Australia or the Ministry of Health.

Contact Details
Health Services Assessment Collaboration (HSAC)
Health Sciences Centre
University of Canterbury
Private Bag 4800
Christchurch 8140
New Zealand
Tel: +64 3 345 8147  Fax: +64 3 345 8191

Email: hsac@canterbury.ac.nz
Web Site: www.healthsac.net
Executive Summary

Introduction

Screening for colorectal cancer (malignancy of the inside lining of the colon and rectum, referred to as CRC) provides a simple and effective public health intervention to prevent and minimize the impact of this disease on the community. Population over the age of 50 years are at increased risk of CRC and therefore targeted for screening.

New Zealand has high prevalence of CRC, and this is likely to worsen over time with an increasing elderly population. Estimated age-adjusted prevalence of CRC in New Zealand is 43.7 per 100,000 - more than double the comparable world prevalence. Population based surveys suggest that in New Zealand, Māori have lower incidence and lower mortality than non-Māori; however, the incidence is increasing among the Māori and Polynesian population.

Three factors make CRC amenable to population based screening programmes: it has slow progression in its natural history, simple screening procedures are available for its detection, and it is curable when detected in early stages (WHO, 2009). Firstly, CRC is a slow growing tumour with a long preclinical phase. Significant Adenoma (tumours that are either more than 10 mm in size, or where histological examination shows more than 20% villous adenomatous tissue, abbreviated in this report as SA) become CRCs at a rate of roughly 1% per year, and if left in situ have a cumulative risk of about 24% of becoming malignant at 20 years. Further, the development of invasive cancer from a small adenoma (<10 mm) is extremely unlikely in less than five years (Allison, 2003). This review therefore will consider CRC and SA together (referred to hereafter as advanced neoplasms). Secondly, effective screening procedures are available for early diagnosis and screening of advanced neoplasms (CRC and Significant Adenoma). Thirdly, CRC is resectable and curable when found in its early stages.

A systematic review of results from four randomised controlled trials of Hemoccult screening (a type of standard guaiac Faecal Occult Blood Test (FOBT) based screening test, explained later in the document) showed that those allocated to screening had a reduction in mortality from CRC of 16% (relative risk 0.84 (95% confidence interval 0.77 to 0.93)). The effect was stronger by about 7% when the results were adjusted for attendance for screening programmes; this reduction was 23% (relative risk 0.77 (0.57 to 0.89)) for people actually screened (Towler et al., 1998). It is believed that establishment of early treatment can reduce mortality of people with CRC. Screening is a way of identifying asymptomatic individuals who have early stage CRC or adenomas. Given these considerations, there has been increasing interest in the introduction of a national programme to screen asymptomatic New Zealanders in order to reduce the burden of this disease (Parry et al., 2007; Working Party on Screening for Colorectal Cancer, 1998).

Faecal Occult Blood Test (FOBT) is commonly used for screening for CRC. FOBT belong to a family of non-invasive stool based tests that detect the presence of peroxidase in the specimens of stool collected from an individual during screening. Peroxidase activity is found in both human and animal blood and in some vegetables.
and fruits. Since large adenomas and cancers tend to bleed, presence of blood in stool is considered to be a possible marker of the presence of either precancerous colonic adenoma or malignancy.

Two common types of FOBT are:

(i) Guaiac Based FOBT (referred to as GT)

(ii) Immunochemical FOBT (iFOBT), or Faecal Immunochemical Test (referred to as FIT).

Guaiac based FOBT (referred to here as GT; alternative names: GFOBT, gFOBT, or cFOBT) detects the peroxidase activity found in haemoglobin when it interacts with a guaiac-impregnated card in the presence of a hydrogen peroxide developer. A positive result is indicated by the immediate appearance of a blue colour on addition of the hydrogen peroxide developer.

There are two types of guaiac based FOBT: standard, and sensitive GT. Both standard and sensitive GT use developers to identify peroxidases in stool; however, sensitive GT can detect lower levels of peroxidase activity than those detected by standard GT. Because both tests are based on their ability to detect peroxidase activity in stool, they can also detect the presence of peroxidases that have originated from other sources. Dietary substances can result in false positive (e.g. rare red meat, turnips, horseradish) or false negative (e.g. vitamin C) results. As a result, both standard and sensitive GT require dietary restrictions before the tests are conducted. However, laboratory based evidence and clinical data suggest that most dietary limitations can be eliminated by waiting for three days before faecal occult blood test development (Sinatra, St. John, & Young, 1999); (Paul Rozen, Knaani, & Samuel, 1999). A recent systematic review found that requirement of restriction in diet for guaiac based FOBTs was associated with reduced patient compliance (Walsh & Terdiman, 2003).

Immunochemical FOBT (FIT) detects the presence of globin in human haemoglobin by direct immunochemical reactions. The tests require sample collection from one, two, or three specimens of stools. Since FIT is chemically specific for bleeding that is limited to the colon and rectum and bleeding of human origin, dietary restrictions are not necessary.

Fibreoptic colonoscopy is the criterion standard for evaluation of a positive FOBT, and is used to diagnose advanced neoplasms. This is an endoscopic procedure to visualize presence of significant adenoma and CRC throughout the length of colonic mucous membrane. The procedure is generally performed using a 160-cm flexible endoscope. Extensive training is required to perform the procedure safely and effectively. Pre-operatively, the patient must have only consumed clear liquids and then consume some form of purgative (low-volume sodium phosphate purge or high-volume polyethylene glycol purge). The patient frequently does not experience pain during the examination, and cannot recall the procedure.

Colonoscopy is generally safe under expert hands. Perforation of the colon is the most serious short term post operative complication and carries an overall risk of 1 in 500 procedures; for screening patients, the risk is less than 1 in 1,000 (Rex et al., 2006). Other potential complications include bleeding, and a 1 in 10,000 risk of death (Adler
et al., 2007; Allison & Lawson, 2006; Atkin, 2003; Bazensky, Shoobridge-Moran, & Yoder, 2007; Bowles et al., 2004).

It is unclear which of standard GT, sensitive GT, and FIT is best suited for CRC screening. Advantages of GT based tests over FIT include lower costs, easier deployment in clinics by physicians, and lower overall positivity rates in turn requiring fewer colonoscopies. Sensitive GTs are often as sensitive as FIT and have higher sensitivity than standard GTs but they have lower specificities than standard GT for the diagnosis of CRC. On the other hand, FIT has higher screening sensitivity than a standard GT, perhaps similar sensitivity as a sensitive GT, and higher specificity compared to both standard and sensitive GT for CRC. Unlike GT, FIT does not require dietary restrictions before testing. Guaiac based tests are believed to be generally best at detecting large, more distal lesions (Burch et al., 2007). Since FIT is more sensitive than standard GT, it may have higher test positivity thus requiring a larger number of positive individuals to be followed up with colonoscopy. This implies more resources for organizing colonoscopy will be required for a FIT based screening programme. Because FIT is a quantitative test, the positivity rates of FIT can be altered by varying the threshold of detection of haemoglobin and can be matched to the best performing guaiac based test. FIT requires fewer samples to be tested and the method can be automated; in comparison, being a qualitative test, both standard and sensitive GT are open to errors due to biases from testers and observers (Kerr, Broadstock, Day, & Hogan, 2007; Kerr, Day, Broadstock, Weir, & Bidwell, 2007; L. Rabeneck et al., 2007; Linda Rabeneck, Zwaal, Goodman, Mai, & Zamkanei, 2008). Performance of standard or sensitive GT depends on who reads the test results: physicians versus lesser trained technicians leading to variations in sensitivity of the test. The following table summarises these issues (Table 1).

Table 1: A summary of comparisons of features between sensitive GT, standard GT, and FIT

<table>
<thead>
<tr>
<th>Different Types of GT</th>
<th>Different Types of FIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two types of GT being used are standard GT (Hemoccult or Hemoccult II) and sensitive GT (Hemoccult Sensa or Hemoccult II Sensa). Sensitive GT has more ethanol and other components than standard GT that make it able to detect lower levels of peroxidase in stool.</td>
<td>The differences in FITs are related to the way they bind to the globin part of the haemoglobin and the types of sampling required. These include (1) brush based sampling, (2) stick based sampling, and (3) probe based sampling.</td>
</tr>
<tr>
<td>Sensitive GT is more sensitive than standard GT, but has lower specificity and higher positivity rates.</td>
<td>FIT has higher positivity rates than GT; FIT is more sensitive than standard GT but not more sensitive than sensitive GT; FIT can be more specific than sensitive GT.</td>
</tr>
<tr>
<td>Subject to observer bias since the tests have to be read by physicians or technicians.</td>
<td>The process is automated.</td>
</tr>
<tr>
<td>Both standard GT and sensitive GT require dietary restrictions; however, if development of samples for analysis is delayed beyond 48 hours, the need for dietary restrictions may be bypassed. This will in turn make the waiting time for results longer.</td>
<td>FIT does not require any dietary restriction.</td>
</tr>
</tbody>
</table>

A recent Cochrane Systematic Review of CRC screening using the Hemoccult Faecal Occult Blood Test (standard GT) found that screening had a 16% reduction in mortality from CRC (Hewitson, Glasziou, Watson, Towler, & Irwig, 2008). It is therefore believed that if another FOBT can be identified that has better screening...
performance than standard GT, use of this test might be associated with an even greater reduction in mortality from CRC (Allison & Allison, 2005; Allison & Lawson, 2006).

The purpose of this review is to compare screening properties of standard GT, sensitive GT, and FIT for the detection of advanced neoplasms. This comparison will be based on the results of primary studies and secondary data analyses identified in the literature.

Specifically, this review will:

(1) Conduct a literature review to compare sensitivity, specificity, positive predictive value and number of colonoscopies needed to identify one neoplasm for standard GT, sensitive GT, and FIT; and

(2) On the basis of this review data, attempt to identify which screening test might be the best choice for a screening programme in terms of screening properties and selected parameters.

**Materials and methods**

HSAC conducted a systematic review of the literature on published data about screening tests for CRC. The steps were as follows:

(1) Step One – Framing the research question: A research question was formulated using the Patient-Intervention-Comparator-Outcome (PICO) framework. Using the specified search terms, bibliographic databases were mined to identify current research on direct comparison of GT and FIT (limited to preceding five years, i.e. January 2003- December 2008 and studies published in English language or with appropriate translations that were already available).

(2) Step Two – Evaluation of Titles/Abstracts: The retrieved studies were initially evaluated based on titles and abstracts as a first pass to identify studies to be retained for further analysis, using the PICO criteria mentioned above.

(3) Step Three – In depth analysis of individual studies: Full-text versions of selected publications were critically appraised with respect to the following criteria: (a) quality of the studies (nature of the research), (b) study population, (c) type of FOBT and criterion used for validating the FOBT, (d) comparator screening programme, and (e) the health outcomes (Significant Adenoma, and CRC). Reference lists were scanned for each article to search for additional publications that could be further reviewed for this study.

(4) Step Four – Data analysis: For each article, following critical appraisal the results were abstracted from the data presented in the publication into a database for further analysis. The main results and conclusion section of the selected and appraised articles were summarized and data from these articles were then further analysed by pooling together the data wherever indicated.

**Key results**

The search strategy identified a total of 97 citations. After consideration of titles and abstracts using the study selection criteria, a set of seven full text publications were retrieved and scrutinised in detail for possible inclusion in the review.
Three systematic reviews of the literature were identified that compared performances of different brands of GT versus FIT. The review by Burch et al (2007) identified 59 studies – 10 out of 59 studies compared head to head either between sensitive GT and FIT, or between standard GT and FIT. This study reported results of two cohort studies and eight case control studies.

Of the two diagnostic cohort studies, one reported that Hemoccult Sensa (sensitive GT) was less sensitive and less specific than Flexsure (FIT) for the detection of CRC. The other diagnostic cohort study found Hemoccult Sensa (sensitive GT) was the most sensitive but least specific for the diagnosis of CRC. In that study, performance of Immudia HemeSp (FIT) was intermediate: it was less sensitive and more specific than Hemoccult Sensa (sensitive GT), but more sensitive and less specific than unrehydrated Hemoccult (standard GT). Hemoccult (standard GT) had the lowest sensitivity and highest specificity.

Out of the eight case control studies, four used all neoplasms as outcomes (they did not indicate specific size of tumours; presumably “all neoplasms” indicated all advanced adenomas). In these four studies, it was not mentioned whether they compared specifically sensitive or standard GT versus FIT: sensitive and standard GT were treated as GTs. However, it was not possible to infer from the results from this systematic review whether the performance of FIT was superior to either standard or sensitive GT, for the detection of CRC. To summarize the results of this review: (1) for the detection of all neoplasms, three out of four studies reported that FIT had higher sensitivities compared to GTs, and (2) for the detection of CRC, two studies found FIT to have higher sensitivity and two found GT to have higher sensitivity. However they found that overall, Immudia HemeSp (FIT) had superior sensitivity and specificity profiles, averaging 62.6% sensitivity and about 94.3% specificity for the detection of CRC, and these figures were better than either sensitive or standard GT evaluated in the studies included in their review.

Researchers at the Technology Evaluation Centre of the Blue Cross Blue Shield in the United States (2004) performed a comprehensive literature review, but did not post any sensitivity or specificity figures. They found insufficient evidence relevant to a US context to draw any conclusion whether one type of FOBT was superior to another in the detection of either advanced neoplasms or CRC.

In the third review, researchers at the Medical Services Advisory Committee of Australia (2004) conducted a review comparing performances of the different types of FIT and GT for the detection of advanced neoplasm and CRC. They identified nine overall comparisons but three studies reported head to head GT versus FIT performances in terms of reported sensitivities and specificities. In summary, they found that in two of these comparisons, Hemeselect (FIT) was more sensitive than Hemoccult (standard GT) or Hemoccult Sensa (sensitive GT). Both Hemoccult and Hemoccult sensa were more specific for significant adenoma or CRC compared to Hemeselect (FIT). Also, Hemoccult (standard GT) was both more sensitive and specific when compared with Fecatwin (FIT) for advanced neoplasms. The authors concluded that overall, evidence was inconclusive whether FIT (considering both Fecatwin and Hemeselect) was better than either standard or sensitive GT for screening of advanced neoplasms. In this review, they did not differentiate between standard and sensitive GT) for the diagnosis and screening of advanced neoplasms on
the basis of three studies. Further, they commented that all these studies were open to verification bias because of the time lag between screening and diagnosis.

In summary, based on the analysis of systematic reviews, there is inconclusive evidence whether one or other brand of FIT (Fecatwin, Flexsure, ImmodiaSp, or Hemeselect) has a clear advantage in terms of screening performance over any of either standard GT (Hemoccult) or sensitive GT (Hemoccult Sensa) for the diagnosis of advanced neoplasms. However, these studies also suggest that FIT might be more sensitive but less specific than either standard GT or sensitive GT for diagnosis of CRC (Burch, et al., 2007; Medical Services Advisory Committee, 2004; Piper, 2004).

HSAC identified one large RCT conducted on an average risk population in the Netherlands (N = 20,623) by van Rossum et al. (2008). In this study, 6,197 individuals tested with OC-Sensor (FIT) were compared with 4,638 individuals who were tested with Hemoccult II (standard GT). Colonoscopy was used as the criterion standard for the diagnosis of advanced neoplasms and was conducted only on positive individuals. Thus the sensitivity could not be calculated, but the authors reported specificity, positive predictive value (PPV), positivity, and detection rate for advanced neoplasms. The authors concluded that standard GT as a screening tool for average risk adults for CRC was likely to lower estimation of the prevalence of CRC (van Rossum et al., 2008).

HSAC identified three observational epidemiological studies that compared different FIT with different types of GT. In a large prospective study, Allison et al. (2007) studied 5841 average risk individuals in the United States to evaluate the performance of Flexsure OBT (FIT) compared with Hemoccult Sensa (sensitive GT), as well as the performance of both tests taken together for advanced neoplasms and CRC. For advanced neoplasms, they found that Hemoccult Sensa (sensitive GT) had higher sensitivity at 43.1% (95% CI: 34.7%-51.8%) than Flexsure (FIT) with sensitivity figures of 33.1% (95% CI: 24.9%-42.3%). However, Hemoccult Sensa was less specific for these groups of lesions at 90.7% (95% CI: 89.9%-91.5%) compared to Flexsure at 97.5% (95% CI: 97%-97.9%). It is important to note that Flexsure (FIT) had significantly higher PPV 23.1% (17.2%-30.3%) than Hemoccult Sensa at 10.1% (7.1%-12.9%). This study supported the finding that Flexsure (FIT) had slightly lower sensitivity but significantly higher specificity compared to a sensitive GT, at the same time required fewer colonoscopies to confirm one case of advanced neoplasm (Allison et al., 2007).

Alicia Smith and colleagues (2006) conducted pairwise comparison of positivity rates between Insure (FIT) and Hemoccult Sensa (sensitive GT) in 2351 screening cohorts and 161 symptomatic individuals (diagnostic cohorts) in Australia. For the screening cohort alone, they found FIT had higher sensitivity but lower specificity than the sensitive GT for advanced neoplasms. They commented that FIT was significantly better at detecting advanced neoplasms; the true positive rate (TPR) for advanced neoplasms was higher with FIT than sensitive GT (44.4% vs. 24.2%), and PPV for FIT and sensitive GT for advanced neoplasms were 26% and 20.2% respectively. They concluded that the brush-sampling Insure (FIT) was more sensitive for advanced neoplasms than a Hemoccult sensa – a sensitive GT (Smith, Young, Cole, & Bampton, 2006).
In an Israeli population, Rozen and colleagues compared OC-Micro (FIT) with Hemoccult II Sena (sensitive GT). In this analysis, they varied the threshold of detection of haemoglobin in steps of 25 ng/ml, stepping up the threshold values of the FIT from low 25 ng/ml to 200 ng/ml for OC-Micro, and at each threshold value, compared the screening parameter values of OC-micro with Hemoccult Sena. For threshold set at detecting haemoglobin at levels of 125 ng/100 ml faeces, they found that OC-micro (FIT) had 75% sensitivity while Hemoccult sensa (sensitive GT) had 53.1% sensitivity. At this threshold level, OC-micro had 86.2% specificity of 86.2% and Hemoccult sensa 59.4% specificity for advanced neoplasms. At this threshold of detection, compared to 12.1% PPV for sensitive GT for advanced neoplasms in this study, FIT had 38.1% PPV. This indicated that FIT had lower number of colonoscopies needed to confirm one case of advanced neoplasm than sensitive GT. Using Hemoccult sensa (sensitive GT) about eight colonoscopies would be required for positive patients to confirm one case of advanced neoplasm; using OC-Micro would require about two colonoscopies to confirm one case of advanced neoplasm (P Rozen et al., 2008).

Results from all three observational epidemiological studies found that FIT in general had higher PPV than the Hemoccult Sena brand of GT. This indicated that while FIT may have higher positivity rates compared to either standard or sensitive GT for the diagnosis of advanced neoplasms, FIT might also be associated with more efficient use of colonoscopy services for the detection of advanced neoplasms as fewer colonoscopies would be required for the identification of one neoplasm, leading to better utilisation of colonoscopy although overall rates of colonoscopy would likely to be higher.

In summary, (1) systematic reviews found for advanced neoplasm, FIT had similar sensitivity and specificity as sensitive GT; however, FIT was more sensitive and less specific for CRC; overall, there was insufficient evidence to conclude if FIT was more appropriate than sensitive or standard GT as a screening alternative; (2) the randomized controlled trial found similar sensitivity, specificity, PPV (and therefore numbers needed to detect, abbreviated as NND, also known as numbers needed to scope or NNS) between FIT and sensitive GT for advanced neoplasms; (3) observational epidemiological studies identified that brands of FIT (Flexsure, Insure and OC-Micro) had higher positivity rates, higher specificities, and therefore higher PPV estimates, and correspondingly lower NND estimates than either standard or sensitive GT for the detection of advanced neoplasms and CRC. The results from observational epidemiological studies indicated that the use of FIT compared to either standard GT or sensitive GT might lead to more colonoscopies, but at the same time fewer colonoscopies would be required to confirm one case of advanced neoplasm, and thus more efficient use of colonoscopy services for the detection of advanced neoplasms. These findings suggested that if the threshold levels of detecting Hb were set at 100 ng/ml of faeces (which is roughly equivalent to detecting 30 ng Hb/100g faces), use of any of the three commonly used brands of FIT (Insure, Flexsure or OC-Micro) would have comparable sensitivity, higher specificity, and higher PPV and lower NND than Hemoccult Sena (sensitive GT) for the detection of advanced neoplasms. In essence, compared to the use of sensitive GT, use of FIT would likely to be associated with increased positivity, but increased requirement of colonoscopy testing might also lead to more efficient use of colonoscopies as fewer colonoscopies are required for FIT than sensitive GT for the detection of advanced neoplasms.
In addition to critical appraisal of primary studies and the literature reviews, data from the primary studies were pooled to compare the performances of FIT versus sensitive or standard GT. The first and third quartiles of sensitivity, specificity, positive predictive value, and numbers needed to detect (inverse of positive predictive value; also known as numbers needed to scope and indicates how many colonoscopies would be required to detect one confirmed case of CRC or significant adenoma; abbreviated as NND) for each of FIT and standard or sensitive GT were compared for advanced neoplasms. For sensitive GT, the median sensitivity was 50% (IQR 43.1% - 53.1%), the median specificity was 90.7% (IQR 80.1% - 97.1%), the median PPV was 20.2% (IQR 12.7% - 55.2%) and the corresponding median NND was 4.95 (1.81, 7.86). For FIT, the median sensitivity was 67.5% (IQR 42.6% - 73.7%), the median specificity was 91.1% (IQR 80.1% - 97.7%), the PPV was 38.1% (28.2% - 54.1%), and the corresponding NND was 2.62 (1.86 - 3.53). These figures suggest that FIT might have higher sensitivity, specificity, and PPV (correspondingly lower NND) values for advanced neoplasms compared to sensitive GT. However the figures are suggestive of median estimates and were obtained from heterogeneous studies. Therefore, this interpretation requires caution and further investigation.

Conclusions

In summary, for advanced neoplasms, systematic reviews of direct comparison between FIT and standard or sensitive GT did not find sufficient evidence to state whether one FOBT is preferable to another; however for CRC, FIT had higher sensitivity and specificity than standard or sensitive GT. HSAC identified only one RCT, but this study reported similar figures for sensitivity, specificity and PPV between OC-Sensor (FIT) and Hemoccult Sensa (sensitive GT). HSAC identified three observational epidemiological studies with three different brands of FIT (Flexsure, Insure and OC-Micro) found that while FIT might have lower sensitivity compared to sensitive GT (Hemoccult Senza or Hemoccult II Sensa), all three studies found that at commonly used threshold values, FIT had higher specificity, and higher PPV compared to sensitive GT. When the data of all the tests were pooled together, FIT as a class of test had higher PPV and therefore lower NND compared to sensitive or standard GT for the diagnosis and screening of advanced neoplasms.

These suggest that although use of FIT might lead to a higher percentage of individuals being identified as positive for advanced neoplasms, because this family of tests also had lower NND compared to the family of sensitive GT tests in the primary observational studies, this suggests that FIT – when considered only on the basis of screening parameters – might result in more efficient use of colonoscopy to confirm cases of advanced neoplasms.

However, the findings of this review need to be interpreted in the light of several limitations and restrictions of the approach taken by this review. First of all, this review is based on very few studies to justify the conclusions. Second, this review has limited itself to consider only screening parameters while conducting head to head comparisons between standard & sensitive GT and FITs. These include sensitivity, specificity, positivity, positive predictive value, and number of colonoscopies needed to confirm one case of advanced neoplasm (NND or NNS). None of these measures by itself can provide any guidance as to the superiority of a screening test for a given diagnosis. A sensitive GT is more sensitive than standard GT but is also less specific.
This indicates while a sensitive GT is less likely to miss true positive cases, being low in specificity, it leads to higher false positive individuals who nevertheless need to be followed up by colonoscopy. Thus, a test that is both highly sensitive and highly specific might lead to a balance between increased positivity, increased utilization of colonoscopy services as well as fewer colonoscopies to detect one true case of CRC or advanced neoplasm, leading to better utilization of these services and therefore desirable. Since the thresholds of detection of FIT can be varied to adjust the optimum sensitivity and specificity is advantageous. Third, other factors regarding choice of screening tests must be considered before their deployment but these were not addressed as part of this review. These include relative costs of the screening tests, difficulty of procedures and options that might be acceptable to the population, compliance of, uptake or adoption by people to whom the screening options are offered, and in the case of CRC, issues around preparation of the screening population and interpretation of the results. Fourth, as stated above, this review has attempted to evaluate two rival screening options (guaiac based tests versus faecal immunochemical tests) in terms of their screening performances only. It did not identify studies that had compared sensitive guaiac versus standard guaiac, nor did it identify studies that had compared different types of FIT tests. Only one study in this review had data about comparison of different threshold levels of FIT (OC-Micro) with each other and with GT (Hemoccult II Sensa). In this study by Rozen and colleagues in Israel (2008), OC-Micro was found to have lower test sensitivity at levels of 125 mg/100 g of faeces (equivalent of ng/ml), but even at these thresholds, the specificity of OC-Micro (FIT) was better than that of Hemoccult II Sensa, as was the corresponding PPV, indicating that even when the thresholds of detection are set at higher levels than usual, OC-Micro could still be better in terms of specificity, PPV and NND. Therefore, while this review found some evidence in support of superior performance of FIT tests as a family of screening options over sensitive guaiac tests (and some indirect evidence of superiority of sensitive guaiac tests over standard guaiac tests because of their designed increased sensitivity to identify pseudoperoxidase activity in blood at lower concentrations), the results of the study cannot provide any indication whether one form of FIT is superior to another. Other, newer forms of FOBT and other types of screening tests were not considered as part of this review. Finally, the time frame for this review was between 2004-2008. This time frame was selected because comprehensive systematic reviews based on research on CRC screening prior to 2004 were already available and were included in this review. This review found some support from observational epidemiological studies that different forms of FIT (Flexsure, OC-Micro, and Insure) were superior to GT for screening of CRC.

In conclusion, this review has evaluated the relative screening properties of standard and sensitive guaiac versus faecal immunochemical tests as possible screening options for advanced neoplasms, in order to inform a process of selection of the better of the two tools for screening of CRC. While the systematic reviews and RCTs were inconclusive whether FIT or GT were likely to be better alternatives, observational epidemiological studies found that even with low sensitivity, all three brands of FIT (Flexsure, Insure and OC-Micro) were superior in PPV and NND when compared to Hemoccult Sensa or Hemoccult II Sensa. This indicates that performance wise, as a single test alternative, FIT might be deemed to make better use of colonoscopy resources, even though they may result in higher overall positivity.
# Table of Contents

**Review Team** ................................................................................................................. i

**Acknowledgements** ....................................................................................................... i

**Copyright Statement & Disclaimer** ............................................................................. i

**Contact Details** ............................................................................................................. ii

**Executive Summary** .................................................................................................... iii

  - Introduction .................................................................................................................... iii
  - Materials and methods ..................................................................................................... vi
  - Key results ........................................................................................................................ vi
  - Conclusions ..................................................................................................................... x

**Table of Contents** ....................................................................................................... xii

**List of Tables** ............................................................................................................. xiv

**List of Abbreviations and Acronyms** ....................................................................... xv

**Introduction** .................................................................................................................. 1

  - Objective ......................................................................................................................... 3
  - Structure of report ............................................................................................................ 4

**Methods** ......................................................................................................................... 5

  - Research questions .......................................................................................................... 5
  - Literature search ............................................................................................................... 5
  - Bibliographic databases .................................................................................................... 7
  - Review databases .......................................................................................................... 7
  - HTA websites .................................................................................................................. 7
  - Guideline websites .......................................................................................................... 7
  - Inclusion and exclusion criteria ........................................................................................ 7
  - Assessment of study eligibility ......................................................................................... 8
  - Quality Assurance procedures ........................................................................................ 9
  - Appraisal of included studies .......................................................................................... 9
  - Dimensions of evidence ................................................................................................. 9
  - Data extraction ............................................................................................................... 10
  - Data abstraction .............................................................................................................. 14
  - Data synthesis & reporting ............................................................................................. 14
  - Limitations of the review methodology .......................................................................... 14

**Results** .......................................................................................................................... 17

  - Evidence from reviews and primary studies ............................................................. 17
  - Evidence from systematic reviews ............................................................................ 17
  - Evidence from randomized controlled trials ............................................................. 20
  - Evidence from observational epidemiological studies ............................................ 23
  - Results from simple pooled data analysis ................................................................... 26
Summary and Conclusions .................................................................33
Summary of evidence for evidence review ........................................33
Discussion and conclusions .............................................................33
References ..................................................................................35
Glossary ....................................................................................39
Appendix A: Included Studies ..........................................................47
Appendix B: Excluded Studies .........................................................49
Appendix C: Quality Checklists for Appraising Interventions ..........57
Appendix D: Performances at Various Threshold Levels of Faecal
Immunochromeical Tests ...............................................................59
Comparison of the performances of Hemoccult Sena at various threshold
levels versus Faecal Immunochromeical Tests on the basis of efficiency, safety,
impact and feasibility ...............................................................59
# List of Tables

| Table 1: | A summary of comparisons of features between sensitive, standard GT, and FIT ..........................................................v |
| Table 2: | Criteria for determining study eligibility .............................................7 |
| Table 3: | Nature of the evidence ...........................................................................8 |
| Table 4: | Reasons for exclusion.............................................................................9 |
| Table 5: | Dimensions of evidence based on Broadstock (2000)..............................10 |
| Table 6: | The measures of screening effectiveness of different colorectal screening tests (GT and FIT) ..............................................................11 |
| Table 7: | Table of key terms................................................................................12 |
| Table 8: | Reporting biases in systematic reviews* .............................................16 |
| Table 9: | Results from systematic reviews...........................................................19 |
| Table 10: | Summary tables from the study by Burch et al for Guaiac Tests ...20 |
| Table 11: | Summary table from the study by Burch et al for Faecal Immunochemical Tests.................................................................20 |
| Table 12: | Summary table from the study by Burch et al for direct comparisons for advanced neoplasms.........................................................20 |
| Table 13: | Summary table from the abstracted results of the study by the Australian MSAC for the diagnosis of advanced neoplasms..........20 |
| Table 14: | Summary of the key findings from the van Rossum Study.................22 |
| Table 15: | Abstracted data from the study by van Rossum et al for the diagnosis of advanced neoplasms .................................................................24 |
| Table 16: | Abstracted data from the study by Allison et al.................................24 |
| Table 17: | Abstracted data from the study by Smith et al.................................24 |
| Table 18: | Abstracted data showing effect of varying the threshold for FIT values, in this case OC-Micro, from the study by Rozen et al .........25 |
| Table 19: | Summary of key studies and key results..................................................27 |
| Table 20: | Comparison between GT and FIT for significant adenoma and CRC for sensitivity, specificity, and NND ........................................30 |
| Table 21: | Pooled data from the primary studies..................................................31 |
| Table 22: | Comparison of sensitive GT with faecal immunochemical tests for different threshold levels of detection for the faecal immunochemical test .................................................................60 |
List of Abbreviations and Acronyms

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRC</td>
<td>Colorectal Cancer</td>
</tr>
<tr>
<td>DOR</td>
<td>Diagnostic Odds Ratio</td>
</tr>
<tr>
<td>FIT</td>
<td>Faecal Immunochemical Test</td>
</tr>
<tr>
<td>FN</td>
<td>False Negative</td>
</tr>
<tr>
<td>FOBT</td>
<td>Faecal Occult Blood Tests</td>
</tr>
<tr>
<td>FP</td>
<td>False Positive</td>
</tr>
<tr>
<td>GT</td>
<td>Guaiac Test</td>
</tr>
<tr>
<td>HOS</td>
<td>Hemoccult Sensa</td>
</tr>
<tr>
<td>HSAC</td>
<td>Health Services Assessment Collaboration</td>
</tr>
<tr>
<td>IQR</td>
<td>Interquartile range, or the range of numbers between 25th and 75th percentile</td>
</tr>
<tr>
<td>LR</td>
<td>Likelihood Ratio</td>
</tr>
<tr>
<td>MSAC</td>
<td>Medical Services Advisory Committee</td>
</tr>
<tr>
<td>N</td>
<td>Number</td>
</tr>
<tr>
<td>NHS</td>
<td>National Health Service</td>
</tr>
<tr>
<td>NND</td>
<td>Numbers needed to detect (same as NNS)</td>
</tr>
<tr>
<td>NNS</td>
<td>Numbers needed to scope</td>
</tr>
<tr>
<td>PPV</td>
<td>Positive Predictive Value</td>
</tr>
<tr>
<td>RFP</td>
<td>Relative False Positivity</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver operating characteristic curve</td>
</tr>
<tr>
<td>RSN</td>
<td>Relative Sensitivity</td>
</tr>
<tr>
<td>SA</td>
<td>Significant Adenoma</td>
</tr>
<tr>
<td>SROC</td>
<td>Summary receiver operating characteristic curve</td>
</tr>
<tr>
<td>TN</td>
<td>True Negative</td>
</tr>
<tr>
<td>TP</td>
<td>True Positive</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
</tbody>
</table>
Introduction

Screening for colorectal carcinoma (CRC) provides a simple and effective public health intervention to prevent and minimize the impact of CRC on the community. Colorectal cancer is adenocarcinoma of colon and rectum.

New Zealand ranks amongst the countries with the highest levels of CRC, and the problem is likely to worsen over time with an increasing elderly population. Estimated age-adjusted prevalence of CRC in New Zealand is 43.7 per 100,000, more than double the comparable world prevalence of 20 per 100,000 (Bazensky, et al., 2007). Analysis of 14-year epidemiological survey data (1980-1994) reveals that the annual age-adjusted incidence rates of large intestinal cancer among male and female Māori and male and female Polynesians to be 40%, 40%, 39%, and 29%, respectively, of the total population incidence. However epidemiological data suggest that the incidence rates are increasing for all ethnic groups (Timothy D. Sutton, 1993). Māori have lower mortality rates compared to non-Māori. Cohort studies of the entire New Zealand population (1981-1999) showed CRC mortality trends varied markedly. There were small (10-20%) decreases among non-Māori non-Pacific people, a 50% increase among Māori, and up to 10-fold increase among Pacific people. By 1996-99, all three ethnic groups had similar CRC mortality (Shaw, Blakely, Sarfati, Fawcett, & Peace, 2006).

However, death and disability from CRC can be minimized if asymptomatic individuals deemed to be at risk of developing CRC can be periodically followed up with a population based screening programme. CRC fulfils several criteria of diseases that are amenable to population based screening programmes (WHO, 2009). Firstly, CRC is a slow growing tumour with a long preclinical phase. Significant Adenoma (tumours that are more than 10 mm in size) become CRCs at a rate of roughly 1% per year, and if left in situ, have a cumulative risk of about 24% to become malignant at 20 years. Further, the development of invasive cancer from a small adenoma (<10 mm) is extremely unlikely in less than five years (Allison, 2003). Secondly, effective screening procedures are available for early diagnosis and screening of CRC and Significant Adenoma. Thirdly, the tumour is resectable in the early stages. Systematic review of results from four randomised controlled trials of Hemoccult screening (a form of Faecal Occult Blood Test (FOBT) based screening test, explained later in the document) showed that those allocated to screening had a reduction in mortality from CRC of 16% (relative risk 0.84 (95% confidence interval 0.77 to 0.93)). When adjusted for attendance for screening, this reduction was 23% (relative risk 0.77 (0.57 to 0.89)) for people actually screened (Towler, et al., 1998). It is believed that establishment of early treatment can reduce mortality for people with CRC and that screening is a way of identifying asymptomatic individuals who have early stage CRC or adenomas. Given these considerations, there has been increasing interest in the introduction of a national programme to screen asymptomatic New Zealanders in order to reduce the burden of this disease (Parry, et al., 2007; Working Party on Screening for Colorectal Cancer, 1998).

Faecal Occult Blood Test (FOBT) is commonly used for screening of CRC. Faecal Occult Blood Tests (FOBT) belong to a family of non-invasive stool based tests that detect the presence of haemoglobin in the specimens of stool collected from an individual during screening. Since most adenomas tend to bleed, presence of
unchanged blood in stool is considered to be diagnostic of either precancerous colonic adenoma or malignancy. Two common classes of FOBT are:

(i) Guaiac Based FOBT (referred to as GT, and has two types: Standard GT or Sensitive GT)
(ii) Faecal Immunochemical Tests or immunochemical Faecal Tests (referred to as FIT or iFOBT).

Guaiac based FOBT (referred to here as GT; alternative names: GFOBT, gFOBT, or cFOBT). This family of tests detects the pseudoperoxidase activity found in haemoglobin when it interacts with a guaiac-impregnated card in the presence of a hydrogen peroxide developer. A positive result is indicated by the immediate appearance of a blue colour on addition of the developer.

Two types of guaiac based FOBT are available: standard GT and sensitive GT. Standard GT, as described above, is a test that depends on the detection of pseudoperoxidase. Sensitive GT refer to guaiac tests that allow for the detection of lower levels of peroxidase activity than those detected by standard GT. This test detects lower concentration of pseudoperoxidase activity in stool (brand names Hemoccult Sensa or Hemoccult II Sensa, Fecatwin). As a result, this test is more sensitive and less specific when compared to standard GT.

Since both standard and sensitive GT use peroxidase or pseudoperoxidase activity in stool to detect the presence of blood, they can identify other compounds with pseudoperoxidase activity that is also present in stool. These include dietary substances and can result in false positive (e.g., rare red meat, turnips, horseradish) or false negative (e.g., vitamin C) results. Therefore, both these tests for screening require patients to have dietary restrictions before the test is conducted. However, laboratory and clinical data suggest that most dietary limitations can be eliminated by waiting for 3 days before faecal occult blood test development (Sinatra, et al., 1999); (Paul Rozen, et al., 1999). A recent systematic review found that requirements for restriction in diet for guaiac based FOBTs were associated with reduced patient compliance (Walsh & Terdiman, 2003).

Faecal Immunochemical Tests (FIT) aim to identify globin protein in human haemoglobin by using appropriately designed immunological reactions. The tests are generally more sensitive than either standard or sensitive GTs. The different types of FIT vary in the way samples are collected (using brushes, test tubes, or cards) and require sample collection from either one stool sample (Instant View Immocare), two stool samples (Insure, OC-Hemodia), or three stool samples (Hemeselect, Flexsure OBT now known as iHemoccult, or MonoHaem). HSAC could not identify any study on the association between the number of stool specimens collected and screening performances of individual FITs. FIT can specifically identify bleeding that is limited to the colon and rectum.

Fibreoptic colonoscopy is used to diagnose colorectal neoplasms. This is an endoscopic procedure in which a fibreoptic endoscope is used to visualize presence of Significant Adenoma and CRC throughout the length of colonic mucous membrane. The procedure is generally performed using a 160-cm flexible endoscope. Pre-operatively, the patient must have only consumed clear liquids and then consume some form of purgative (low-volume sodium phosphate purge or high-volume
polyethylene glycol purge). The patient should not experience pain during the examination and often cannot recall the procedure.

Colonoscopy is generally safe under expert hands. The most serious short term post operative complication of fibreoptic colonoscopy is perforation. Epidemiological studies have reported an overall risk of perforation of 1 in 500 but a risk of less than 1 in 1,000 screening patients, primarily because of relatively healthy mucosa in the asymptomatic individuals (Rex, et al., 2006). The other complications include bleeding, and a 1 in 10,000 risk of death (Lieberman et al., 2000); (Adler, et al., 2007; Allison & Lawson, 2006; Atkin, 2003; Bazensky, et al., 2007; Bowles, et al., 2004).

All three FOBT - standard & sensitive GT, as well as FIT - are available as screening tests for CRC. Both standard and sensitive GT are relatively inexpensive and simple compared to FIT tests. FIT is chemically more haemoglobin-specific than either standard or sensitive GT. Both sensitive and standard GT are generally best at detecting large, more distal lesions (Burch, et al., 2007). Sensitive GT is more sensitive than standard GT but less specific (Kerr, Broadstock, et al., 2007; Kerr, Day, et al., 2007; L. Rabeneck, et al., 2007; Linda Rabeneck, et al., 2008). GT is also a qualitative test and therefore open to bias and human reading errors. Because the threshold levels of detection of advanced neoplasms for FIT can be varied, it is possible to adjust FIT in order to match the best properties of either standard or sensitive GT. FIT results can be read by automated analysers, unlike the results of standard or sensitive GT, both of which are open to biases resulting from human interpretation of patterns.

A recent Cochrane Systematic Review of CRC screening using the Hemoccult Faecal Occult Blood Test (GT) has found that screening had a 16% reduction in mortality from CRC (Hewitson, et al., 2008). It is therefore believed that if another FOBT can be identified that has better screening performance compared to GT, use of this test might be associated with even greater reduction in mortality from CRC (Allison & Allison, 2005; Allison & Lawson, 2006).

**Objective**

The purpose of this review is to:

1. Conduct a systematic review of the available literature from published data on a head to head comparison of the screening properties of guaiac based Faecal Occult Blood Test (GT) versus Faecal Immunochemical Tests (FIT). The measures of relative effectiveness considered in this study are:
   (a) Sensitivity
   (b) Specificity
   (c) Positive Predictive Values
   (d) False Positive Rates
   (e) Relative sensitivity Measurements
   (f) Relative False Positivity Ratios
   (g) Detection Rates
   (h) Number of Colonoscopies required per positive test identified in the screening programmes (inverse of Positive Predictive Values, referred to as NND or Numbers Needed to Detect, alternatively NNS or numbers needed to scope)
(i) Overall Screening Positivity
(j) Likelihood Ratios.

2. Based on this head to head comparison, this review has reconstructed data from the reviews to examine whether FIT as a screening tool can be considered more effective and efficient of resources in terms of utilization of colonoscopy based confirmation of diagnosis of CRC and Significant Adenoma.

The first objective has been addressed by using a broad based comprehensive literature review of the head to head comparisons of the two commonly utilised screening tools, and analysis of reconstructed data from the individual selected studies. The second objective of this review has been addressed by statistically analysing the metrics of relative performance indicators of the two tests.

Structure of report

The report is divided into three sections. The first section, Methods, describes the methods and includes the research questions, search strategy, inclusion and exclusion criteria, the data extraction, appraisal and synthesis methods, and the methodological limitations of the evidence review. The Results section considers the included appraised studies, reporting first on the systematic reviews and meta-analyses, and then on the original primary research. Study characteristics and findings are reported in separate tables and synthesised in the text, and the body of evidence for each research question is reported. The final section, Discussion and Conclusion, summarises results, briefly discusses the limitations of the evidence base and identified gaps in knowledge, and presents key conclusions. A Glossary and detailed appendices follow, including all excluded papers annotated by reason for exclusion, and the completed data extraction tables for included papers.
Methods

This was a systematic review of available literature and publications related to head to head comparisons of the relative effectiveness of GT and FIT technologies for the screening of CRC in a population based screening programme that might be applied in the setting of a developed country whose health care system was comparable with that of New Zealand, in terms of quality of health care and the structure or provision of healthcare provided. A systematic process of search, retrieval, selection of studies, analysis of studies and reporting was followed to identify and abstract data from the retrieved studies. These processes are described below.

The sequential steps followed in this review were as follows:

- Framing of research questions
- Searching the literature databases
- Inclusion and exclusion of studies for conducting the review
- Assessment of study eligibility and data abstraction
- Data analysis and summarisation.

Research questions

The primary research questions to be addressed by this review were:

How does FIT as a screening test for average or above average risk populations compare with GT screening tool, when compared in terms of sensitivity, specificity, positive predictive values, False Positive Rates, Relative Sensitivity and Relative False Positive Rates, Detection Rates, Number of Colonoscopies required per positive test identified in the screening programmes, Overall screening Positivity, and Likelihood Ratios.

Based on available data from the literature, what would be the overall summary comparison of performances between GT and FIT as screening tools for the early diagnosis and screening of CRC?

In order to search the relevant research and publication databases, the reviewers framed the following questions to formulate their search strategies. The questions were:

- What are the relative performances of FITs and GTs when used for population health screening?
- How do FITs compare with GT when sensitivity, specificity, PPV, and number of colonoscopies per confirmed diagnosis (NND) are considered?

Literature search

Search terms were formulated based on these questions to capture as much literature on the topic as possible. The search terms included the following terms and concepts associated with Boolean search connectors and wherever the databases allowed, use of fuzzy logic to capture as many relevant publications as possible. The initial search terms, presented here in alphabetical order and in various combinations using Boolean
operators (AND, OR, NOT), fuzzy logic terms and other symbols appropriate for the databases that were searched were as follows:

- Colorectal
- Cancer
- FOBT
- Faecal OR Fecal
- Guaiac
- Immunochemical
- Screening

Searches were limited to English language publications and covered literature that was published in the five preceding years from 2008, the year of the commencement of this review (specifically January 2004-August 2008). This document is based on an earlier limited version of a review where a partial search of the reference lists of key papers and journals were conducted. For the purposes of this initial briefing, the searches were restricted to essential/key bibliographic databases. Websites, abstracts and titles of material identified in the search were scanned electronically for eligibility. Taking the search results retrieved by the previous search, the selected or otherwise eligible studies were then further critically appraised on the basis of quality.

HSAC reviewers searched the World Wide Web (using the popular search engine Google located at http://www.google.com and http://scholar.google.com) and other sources of bibliographic data. The researchers used the following bibliographic databases: Medline, Embase, the Cochrane Database of Systematic Reviews (CDSR), and the Database of Abstracts of Reviews of Effects (DARE) to help identify studies and systematic reviews.

The reference lists of included papers were scanned to identify any peer-reviewed evidence that may have been missed in the literature search. The Web of Science was also searched to identify any relevant papers citing the pivotal references. Hand searching of journals, contacting of manufacturers, or contacting of authors for unpublished research was not undertaken in this review. Whilst grey literature and unpublished material such as conference abstracts were not included in the evidence review, they may be referred to in background sections.

Search terms were searched for as keywords, exploded where possible, and as free text within the title and/or abstract, in the Embase and Medline databases. Variations on these terms were used for Cochrane library and for all other relevant databases were modified to suit their keywords and descriptors.
The databases searched for identifying studies for this review were:

**Bibliographic databases**
- Embase
- Medline

**Review databases**
- Cochrane Database of Systematic Reviews
- Cochrane Central Register of Controlled Trials
- Database of Abstracts of Reviews of Effectiveness
- Health Technology Assessment database
- NHS Economic Evaluation database

**HTA websites**
- INAHTA website database: http://www.inahta.org/Search2/?pub=1
- MSAC: http://www.m sac.gov.au/
- NZHTA: http://nzhta.chmeds.ac.nz/
- NICE: http://www.nice.org.uk/

**Guideline websites**

**Inclusion and exclusion criteria**

Following initial search and retrieval of all possible articles and literature on the topic, an initial appraisal of the literature was performed to filter literature sources that would be retained for full review. For inclusion in the current review, the evidence had to fulfil the criteria outlined in **Table 2** and **Table 3**. These criteria were developed *a priori* and described in the scoping protocol prepared prior to commencement of the review proper.

**Table 2: Criteria for determining study eligibility**

<table>
<thead>
<tr>
<th>Patient population</th>
<th>Population belonging to an age group of 50-75 years, belonging to a developed country or an English speaking country who are at average or above average risk of adenocarcinoma of colon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention</td>
<td>Immunochemical Faecal Occult Blood Test (FIT)</td>
</tr>
<tr>
<td>Comparator</td>
<td>Guaiac Faecal Occult Blood Test (GT)</td>
</tr>
<tr>
<td>Outcomes</td>
<td>Screening performances of each of these tests were compared with the other in terms of the following outcome variables deemed to be representative of screening effectiveness: (a) Sensitivity of the tests, (b) Specificity of the tests, (c) Positive Predictive Value of the tests, (d) Number of colonoscopies needed to establish one True Positive Case of colonic cancer from all positive test subjects, (e) Likelihood Ratios, (f) False Positive Rates and (g) Positivity Rates</td>
</tr>
</tbody>
</table>
A study was retained in the database for full review and data extraction and final analysis if it satisfied the following criteria:

- It was a population based screening study
- It conducted a head to head evaluation of GT with FIT
- The population was comparable and relevant to that in New Zealand
- A full text of the research with an appropriate description of the methods and result of the report was available for review. In cases where the study was conducted in a foreign language, an acceptable standardized English translation was available.

All other studies were excluded, initially at the stage of review of abstracts and titles, then later after review of the full text when that was obtained. A study was not included for abstraction of data and final analysis based on full text if it failed to satisfy all of the above conditions and in addition if it had any of the following conditions:

- The outcome was not CRC or did not include CRC
- The study did not include an appropriate comparison group.

In addition, publications that were neither primary research, nor secondary data analysis in the form of systematic review with clear indications for the selection of data and analytical strategies were also excluded. The reviewers did not include opinion pieces, editorials, letters to the editor and other documents which did not include a clear description of the methods, results and discussions of the implications of the findings.

**Assessment of study eligibility**

Studies were selected for appraisal using a two-stage process. In stage one, titles and abstracts (where available) identified from the search strategy were scanned and excluded as appropriate. In stage two, based on the filter applied in stage one, full texts of articles were retrieved and critically appraised.
Quality Assurance procedures

Each study was assessed for quality of its report based on the following – overall design quality, description of the study design, steps taken to minimize verification bias, steps taken in the study to minimize selection bias, the test itself and its comparison groups, reference standards, and external validity of the study (Medical Services Advisory Committee, 2004).

Initially, a total of 97 non-duplicate studies were identified by the search strategy. As detailed in Table 4, from this initial list seven full text articles were eligible for retrieval after excluding studies from the search titles and abstracts. Reasons are presented hierarchically such that the first reason in the list that applied is reported. Other cited publications (e.g. those providing background materials) are presented in the References (Page 35).

Initial retrieval = 97 articles

**Table 4: Reasons for exclusion**

<table>
<thead>
<tr>
<th>Exclusion Criterion</th>
<th>N</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial retrieval</td>
<td>97</td>
<td>100%</td>
</tr>
<tr>
<td>Inappropriate population on which the study was conducted</td>
<td>5</td>
<td>5.15%</td>
</tr>
<tr>
<td>Study did not include a proper intervention or the intervention was inappropriate for this review (i.e. not related to Faecal Occult Blood Tests in any form either Guaiac or Immunochemical)</td>
<td>44</td>
<td>45.4%</td>
</tr>
<tr>
<td>No comparison between GT and FIT found in the studies</td>
<td>4</td>
<td>4.12%</td>
</tr>
<tr>
<td>Colorectal cancer was not the specific outcome studied in the publication or research</td>
<td>11</td>
<td>11.3%</td>
</tr>
<tr>
<td>Non-epidemiological studies and publication types that were not acceptable for review</td>
<td>26</td>
<td>26.8%</td>
</tr>
<tr>
<td>Final selection from the initial search</td>
<td>7</td>
<td>7.22%</td>
</tr>
</tbody>
</table>

**Appraisal of included studies**

**Dimensions of evidence**

The hierarchy of evidence was appraised as described in the following listing (Table 5). It is important to recognise that the value of a piece of evidence is determined by all of these dimensions, not just the level of evidence.
Table 5: Dimensions of evidence based on Broadstock (2000)

<table>
<thead>
<tr>
<th>Level of Evidence</th>
<th>Characteristics of the Study Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All tests done on each person (within-subjects design)</td>
</tr>
<tr>
<td>2a</td>
<td>Different tests done on randomly allocated individuals (between subjects design, randomized controlled trial)</td>
</tr>
<tr>
<td>2b</td>
<td>Different tests done on randomly allocated groups (between subjects design)</td>
</tr>
<tr>
<td>3a</td>
<td>Different tests done on different individuals, not randomly allocated, and recruited concurrently</td>
</tr>
<tr>
<td>3b</td>
<td>Different tests done on different individuals, not randomly allocated, or random allocation not possible, comparison data taken from historical cohorts</td>
</tr>
</tbody>
</table>

The format presented in Table 5 was generally followed in this review. The results are presented beginning with the results from Systematic Reviews, followed by results from Randomized Controlled Trials and finally, results from Observational Studies.

Data extraction

Data from primary studies and systematic reviews were extracted onto spreadsheets and data tables, which were specifically designed data extraction forms. The data elements considered for this review included information regarding name and affiliation of the first author who reported the results of the study in the publications database, the study design, patient characteristics, description of the screening protocols, relevant outcomes, study quality and relevant results. The list of outcomes included the following:

- Sensitivity of the study
- Specificity of the study
- Positivity of the specific screening protocol (FIT or GT)
- False Positivity Rate
- Positive Predictive Value
- Likelihood Ratio
- Numbers Needed to Detect (NND)
- Relative Sensitivity (RSN)
- Relative False Positivity (RFP)
- Diagnostic Odds Ratio (DOR)
Tables 6 and 7 provide definition and explanation of the terms related to measurement of effects and impacts of screening studies. Table 7 provides definitions of key terms related to performance of a screening test. The criterion standard in all studies considered in this review was colonoscopy.

**Table 6: The measures of screening effectiveness of different colorectal screening tests (GT and FIT)**

<table>
<thead>
<tr>
<th>Decision based on criterion standard diagnosis</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Results</td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td></td>
</tr>
<tr>
<td>Disease Present</td>
<td>TP</td>
</tr>
<tr>
<td>Disease Not Present</td>
<td>FP</td>
</tr>
<tr>
<td>Negative</td>
<td></td>
</tr>
<tr>
<td>False Negative</td>
<td>FN</td>
</tr>
<tr>
<td>True Negative</td>
<td>TN</td>
</tr>
<tr>
<td>Total</td>
<td></td>
</tr>
<tr>
<td>True Positive + False Negative</td>
<td>FN + TN</td>
</tr>
<tr>
<td>False Positive + True Negative</td>
<td>FP + TN</td>
</tr>
<tr>
<td>N</td>
<td></td>
</tr>
</tbody>
</table>

Legend for this table: criterion standard in case of CRC studies is diagnosis by colonoscopy. In one study by Allison et al (2007), participants for whom colonoscopy was not done or was contraindicated were also followed up by flexible sigmoidoscopy. In this study, in addition to colonoscopy, flexible Sigmoidoscopy was the criterion standard of diagnosis.

TP = True Positive, FP = False Positive, FN = False Negative, TN = True Negative, N = Total number of individuals included in the analysis
### Table 7: Table of key terms

<table>
<thead>
<tr>
<th>Screening Criterion</th>
<th>Description</th>
<th>Implication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>Percentage of all persons with the disease identified as positive by the screening test, expressed as True Positive (TP) / sum of True Positives (TP) and False Negatives (FN)</td>
<td>Ability of the test to identify or rule in those with disease</td>
</tr>
<tr>
<td>Specificity</td>
<td>Percentage of all persons who do not have the disease identified as negative for the disease by the test. This is expressed as True Negative / sum of True Negative and False Positives. Under Rare disease assumption as in the case of CRCs, specificity can also be calculated as $1 – FP / (N – TP)$, where FP = false positives, N = total number of individuals who were screened, and TP = True Positives</td>
<td>The ability of the test to rule out those who do not have the disease</td>
</tr>
<tr>
<td>Positive Predictive Value</td>
<td>Percentage of those identified as positive by the test that truly have the disease (TP / (TP + FP))</td>
<td>Ability of the test to predict those who have the disease</td>
</tr>
<tr>
<td>Negative Predictive Value</td>
<td>TN / (TN + FN)</td>
<td>Ability of the test to predict those who do not have the disease</td>
</tr>
<tr>
<td>Pre Test &amp; Post Test Odds</td>
<td>Pre-test Odds = Prevalence / (1 – Prevalence)</td>
<td>What is the likelihood that a person will develop a disease or has a disease, before (Pre Test Odds) and after the test (Post Test Odds) is conducted</td>
</tr>
<tr>
<td>Likelihood Ratio</td>
<td>Sensitivity / (1 – specificity)</td>
<td>The ability of the test to influence the diagnostic decision whether a person has a disease or not</td>
</tr>
<tr>
<td>Relative Sensitivity (RSN)</td>
<td>TP of Test One/ TP of Test Two</td>
<td>As a measure of screening sensitivities this is applicable when the screening reference test cannot be applied to both positive and negative individuals, as in the case of suspected colonic neoplasm, where negative individuals cannot be followed up with colonoscopy for operational reasons</td>
</tr>
<tr>
<td>Positivity</td>
<td>Total number of positive tests relative to the total number of tests returned or total number of persons who were screened and returned test results. In terms of formula, Positivity = TP + FP / N</td>
<td>This indicated how many people would be needed to be followed up with colonoscopy in an actual screening situation in the community</td>
</tr>
</tbody>
</table>
### Table 7: Table of key terms (continued)

<table>
<thead>
<tr>
<th>Screening Criterion</th>
<th>Description</th>
<th>Implication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detection Rate</td>
<td>Per protocol Detection Rate was defined as the number of True Positives identified by the test compared to the total number of persons screened or total number of screenings done or total number of screenings returned by people who were screened. In terms of formula, this is represented as $DR = TP / N$ Some studies have used a concept known as per intention Detection Rate. Here, total number of true positives relative to the total number of people who were invited to participate in the tests was used.</td>
<td>It is a measure of the accuracy of the screening programme</td>
</tr>
<tr>
<td>Numbers Needed to Screen</td>
<td>The inverse of Detection Rate or the ratio of the total number of screening or total number of people who were screened and the number of true positives</td>
<td>The total number of people who are needed to be screened to identify one true case; it is a measure of the efficiency of the screening programme</td>
</tr>
<tr>
<td>Verification Bias</td>
<td>Bias that results because of non-testing of the individuals who screen test negative using the reference standard</td>
<td>In cases of diseases that require a confirmatory test that can be potentially dangerous and harmful for the patients (therefore carcinoma of colon is an example), not all individuals who test negative with the screening test are followed up with the reference standard test. Alternatively, negative individuals may be offered the reference standard tests after many years following initial proof of screen test negativity. This results in an unequal distribution of outcomes detected by the reference standards in the screen positive and screen negative groups</td>
</tr>
<tr>
<td>Number of Colonoscopies needed (NND or numbers needed to detect)</td>
<td>This is the inverse of PPV and indicates the number of colonoscopies needed to identify one true case of colorectal cancer</td>
<td>The formula for this metric is given by: $1/PPV$ The significance of this metric is that the lower the number, the fewer the number of colonoscopies required to identify one case of colorectal carcinoma. This metric can be used in situations where a measure of positivity is unavailable because the studies have conducted colonoscopies only on those individuals who turned out to be positive in one or other test, or in head to head comparisons of those studies where a colonoscopy was done to confirm only the positive tests and no colonoscopy or sigmoidoscopy was offered to individuals who were negative for any of the FOBTs</td>
</tr>
</tbody>
</table>
Data abstraction
HSAC researchers created a spreadsheet for each study included in the review to enable direct comparison of GT and FIT. The abstracted values were entered for analysis.

Data synthesis & reporting
Data Analysis and Presentation of Sensitivity Analyses

A narrative summary of the key feature of each study included in this review is presented below. In addition, HSAC researchers conducted pooled data analysis of the figures abstracted from each of the studies. These are presented separately. The sensitivity analysis is described below.

Pooling of Data

The sensitivities and specificities of selected studies on FIT were pooled together to set up a pooled summary receiver operating characteristic curve (SROC). This was done as follows:

1. For each study in the database for which complete information for the True Positive and False Negative Rates of each procedure (GT and FIT) were available, the True Positive Rate (Sensitivity) and the False Negative Rate (FNR = Total number of False Negatives / (True Positives + False Negatives)) were first calculated
2. For each of True Positive and False Negative Rates, logits were taken of the following entities:
   - Logit of True Positive given by log(TPR/(1-TPR)) referred to as logitTPR
   - Logit of the False Negative Rate given by log(FNR / (1 – FNR)), referred to as logitFNR
3. From 2.1 and 2.2, a linear regression was set up as follows:
   - (logitTPR-logitFNR) = alpha + beta * (logit TPR + logit FNR) + Error Term
4. From Equation 3.1, the estimated values of alpha and beta were applied to solve for the predicted sensitivity of each study given the specificity value of each study

Limitations of the review methodology

The results of this review need to be interpreted in the light of its several biases. Some of these were not addressed or were beyond the scope of this review. Firstly, this review was limited to comparisons of different forms of FOBT (guaiac or immunochemical based). Other, newer forms of FOBT and other types of screening tests were not considered as part of this review. Comparison of different brands of GT or different brands of FIT themselves was also out of the scope of the present review. Secondly, the time frame for this review was between 2004-2008. This time frame was selected because comprehensive systematic reviews based on research on CRC screening prior to 2004 were already available and were included in this review. However, there may have been new research since 2008 that was missed in this review. Thirdly, consideration of screening properties is not the only decisive factor in determining whether one screening test is superior to another for the detection of
Comparison of diagnostic accuracy between immunochemical and guaiac based faecal occult blood tests for colorectal cancer detection

CRC/Significant Adenoma. Other factors that are acknowledged but not addressed in this review include: acceptability of the test to the individuals who are being screened, numbers of tests conducted, and thresholds for FIT tests.

In addition, as a general remark, this review used a structured approach to review the literature. However, there were some inherent limitations with this approach. All types of study are subject to bias, with systematic reviews being subject to the same biases seen in the original studies they include, as well as biases specifically related to the systematic review process. Reporting biases are a particular problem related to systematic reviews and include publication bias, time-lag bias, multiple publication bias, language bias and outcome reporting bias (Egger et al., 2001). A brief summary of the different types of reporting bias is shown in Table 8. Other biases can result if the methodology to be used in a review is not defined a priori (i.e. before the review commences). Detailed knowledge of studies performed in the area of interest may influence the eligibility criteria for inclusion of studies in the review and may therefore result in biased results. For example, studies with more positive results may be preferentially included in a review, thus biasing the results and overestimating treatment effect.
### Table 8: Reporting biases in systematic reviews*

<table>
<thead>
<tr>
<th>Type of bias</th>
<th>Definition and effect on results of review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publication bias</td>
<td>The publication or non-publication of research findings. Small, negative trials tend not to be published and this may lead to an overestimate of results of a review if only published studies are included.</td>
</tr>
<tr>
<td>Time-lag bias</td>
<td>The rapid or delayed publication of research findings. Studies with positive results tend to be published sooner than studies with negative findings and hence results may be overestimated until the negative trials ‘catch up’.</td>
</tr>
<tr>
<td>Multiple publication bias</td>
<td>The multiple or singular publication of research findings. Studies with significant results tend to be published multiple times which increases the chance of duplication of the same data and may bias the results of a review.</td>
</tr>
<tr>
<td>Citation bias</td>
<td>The citation or non-citation of research. Citing of trials in publications is not objective so retrieving studies using this method alone may result in biased results. Unsupported studies tend to be cited often which may also bias results.</td>
</tr>
<tr>
<td>Language bias</td>
<td>The publication of research findings in a particular language. Significant results are more likely to be published in English so a search limited to English-language journals may result in an overestimation of effect.</td>
</tr>
<tr>
<td>Outcome reporting bias</td>
<td>The selective reporting of some outcomes but not others. Outcomes with favourable findings may be reported more. For example, adverse events have been found to be reported more often in unpublished studies. This may result in more favourable results for published studies.</td>
</tr>
<tr>
<td>Verification bias</td>
<td>Bias in screening studies that arise because of long time lag between screening and confirmatory tests</td>
</tr>
</tbody>
</table>

* Adapted from Egger et al. (2001).

Some of these biases are potentially present in this review. Only data published in peer-reviewed journals is included. No attempt was made to include unpublished material, as such material typically has insufficient information upon which to base quality assessment, and it has not been subject to the scrutiny of the peer-review process. In addition, the search was limited to English-language publications so language bias is also a potential problem. Outcome reporting bias and inclusion criteria bias are unlikely as the reviewers had no detailed knowledge of the topic literature, and the methodology used in the review and scope of the review was defined *a priori*.

The review was conducted over a limited timeframe (December 2008 – April 2009).

For a detailed description of interventions and evaluation methods, and results used in the studies appraised, the reader is referred to the original papers cited.
Results

Evidence from reviews and primary studies

The evidence of the head to head performances of the GT versus FIT is presented below based on systematic reviews, randomized controlled trials, and other types of primary observational epidemiological studies. Unless otherwise stated, the outcome of interest in these direct comparisons is the presence of advanced neoplasms, defined as a combination of significant adenoma (SA) and colorectal cancer (CRC).

Evidence from systematic reviews

Three systematic reviews of literature were identified that compared performances of different brands of GT versus FIT. The review by Burch et al. (2007) identified 59 studies – 10 out of 59 studies compared head to head either between sensitive GT and FIT, or between standard GT and FIT. This study reported results of two cohort studies and eight case control studies.

Of two diagnostic cohort studies, one reported that sensitive GT (Hemoccult Sensa) was less sensitive and less specific than FIT (Flexsure) for the detection of CRC, but the other diagnostic cohort study found sensitive GT (Hemoccult Sensa) was the most sensitive but least specific for the diagnosis of CRC. In that study, Immudia HemeSp (FIT) was intermediate: it was less sensitive and more specific than Hemoccult Sensa (sensitive GT), but more sensitive and less specific than standard GT (unrehydrated Hemoccult). Hemoccult (standard GT) had the lowest sensitivity and highest specificity.

Four of the eight case control studies used all neoplasms as outcome diseases (they did not indicate specific size of tumors; presumably all neoplasms indicated all advanced adenomas). In this review, none of the four studies mentioned whether these studies compared specifically sensitive or standard GT versus FIT; sensitive and standard GT were treated as GTs. However, the results from this systematic review were inconclusive since (1) for the detection of all neoplasms, three out of four studies reported that FIT had higher sensitivities compared to GTs, and (2) for the detection of CRC, two studies found FIT to have higher sensitivity and two studies found GT to have higher sensitivity. However they all found that overall, Immudia HemeSp (FIT) had superior sensitivity and specificity profiles, averaging 62.6% sensitivity and about 94.3% specificity and these figures were better than either sensitive or standard GT evaluated in the studies included in their review.

Researchers at the Technology Evaluation Centre of the Blue Cross Blue Shield (2004) in the United States performed a comprehensive literature review, but did not post any sensitivity or specificity figures. They found insufficient evidence that would be relevant to a US context to draw any conclusion whether one form of FOBT was superior to another. In the third comprehensive review, Researchers at the Medical Services Advisory Committee of Australia (2004) conducted a comparative review of the different FIT and GT. In their review, they identified nine overall comparisons but three studies reported head to head GT versus FIT performances in terms of reported sensitivities and specificities. In summary, they found that in two of these comparisons Hemeselect (FIT) was more sensitive than either Hemoccult (standard
GT) or Hemoccult Sensa (sensitive GT). Both Hemoccult and Hemoccult sensa were more specific for Significant Adenoma or CRC compared to Hemeselect (FIT). In addition, Hemoccult (standard GT) was both more sensitive and specific when compared with Fecatwin (FIT) for advanced neoplasms. The authors concluded that this was inconclusive evidence whether FIT (considering both Fecatwin and Hemeselect) was better than GT (either sensitive GT or standard GT - in this review, they did not differentiate between standard and sensitive GT) for the diagnosis and screening of advanced neoplasms on the basis of three studies. Further, they commented that all these studies were open to verification bias because of the time lag between screening and diagnosis (Tables 9-13).

In summary, based on the analysis of systematic reviews, there is inconclusive evidence whether one or other brand of FIT (Fecatwin, Flexsure, ImmudiaSp, or Hemeselect) has a clear advantage in terms of screening performance over any of either standard GT (Hemoccult) or sensitive GT (Hemoccult Sensa). However, FIT tests (Immudia Sp or Hemeselect) were more sensitive than standard GT (Hemoccult) or sensitive GT (Hemoccult Sensa) for advanced neoplasms (Burch, et al., 2007; Medical Services Advisory Committee, 2004; Piper, 2004).
### Table 9: Results from systematic reviews

<table>
<thead>
<tr>
<th>Author &amp; year</th>
<th>Study type and research question</th>
<th>Methods</th>
<th>Key findings</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burch, 2007</td>
<td>Systematic Review of cohort studies and case control studies on diagnostic accuracies comparing GT and FIT tests</td>
<td>Systematic review of eligible studies, quality assurance was done using the QADAS tool.</td>
<td>Reviewed 59 studies, all cohort studies or case control studies. About 23 studies were based on GT, 25 studies on FIT, 10 studies used both types of tests and reported separate results. Overall, this review found that there was little to choose between FIT or GT; studies based on case control studies tended to overestimate diagnostic accuracies compared to those studies based on cohort studies.</td>
<td>There was inconclusive evidence in terms of sensitivity and specificity if FIT was a better screening tool than GT for either colorectal cancer or for all neoplasms.</td>
</tr>
<tr>
<td>Blue Cross Blue Shield, 2004</td>
<td>Systematic review of the diagnostic performance of FIT vs GT</td>
<td>Using well defined criteria for selection of studies, they searched literature databases (not specified which databases were searched) for articles between 1984-2004. Evaluated each study to identify if performances of FIT were superior to GT.</td>
<td>7 studies were identified in this systematic review. Overall, no definite conclusion was provided whether FIT was found to be superior over GT. No summary estimate of performance characteristics was provided.</td>
<td>Inconclusive</td>
</tr>
<tr>
<td>MSAC Review, 2004</td>
<td>Systematic review of the diagnostic and screening effectiveness of different guaiac based or immunochemical based Faecal Occult Blood Tests</td>
<td>This was a systematic review of all studies on the relative effectiveness of different types of guaiac or immunochemical faecal occult blood tests for the screening of colorectal cancers. The authors identified the studies through a systematic search of the literature and using well defined quality assessment criteria, identified studies appropriate for analysis.</td>
<td>The results were based on 9 head to head comparison studies of Faecal Occult Blood Test protocols. There was no clear indication whether immunochemical FOBT was superior to guaiac FOBT on all counts. Briefly, Hemeselect was found to be more sensitive than Hemoccult, while the Hemoccult was more specific than either Hemeselect or Fecatwin Tests (FIT tests); Hemoccult had comparable specificity with Flexsure (FIT test) in high risk individuals; Flexsure OBT (FIT test) had higher specificity compared to Hemoccult Sensa (guaiac FOBT).</td>
<td>Based on this review, it was not clear if one family of tests (guaiac or immunochemical) was clearly superior in all parameters (sensitivity, specificity, therefore likelihood ratios, and positive predictive values). The authors did not specifically mention figures for comparisons between different classes of FOBTs based on their False Positive Rates or Positivity Rates.</td>
</tr>
</tbody>
</table>
Comparison of diagnostic accuracy between immunochemical and guaiac based faecal occult blood tests for colorectal cancer detection

Table 10: Summary tables from the study by Burch et al for Guaiac Tests

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Test</th>
<th>Type</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced neoplasms</td>
<td>Hemoccult</td>
<td>Standard GT</td>
<td>6.2%</td>
<td>98%</td>
</tr>
<tr>
<td>Advanced neoplasms</td>
<td>KryptoHaem</td>
<td>Standard GT</td>
<td>83.3%</td>
<td>98.4%</td>
</tr>
<tr>
<td>Advanced neoplasms</td>
<td>KryptoHaem</td>
<td>Standard GT</td>
<td>47.4%</td>
<td>98.4%</td>
</tr>
<tr>
<td>Advanced neoplasms</td>
<td>Shionogi B</td>
<td>Standard GT</td>
<td>73.4%</td>
<td>60.3%</td>
</tr>
</tbody>
</table>

Table 11: Summary table from the study by Burch et al for Faecal Immunochemical Tests

<table>
<thead>
<tr>
<th>Dx</th>
<th>Test</th>
<th>Type</th>
<th>Sens</th>
<th>Spec</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced neoplasms</td>
<td>OC Light</td>
<td>FIT</td>
<td>5.4%</td>
<td>98.5%</td>
</tr>
<tr>
<td>Advanced neoplasms</td>
<td>Immudia HemSp</td>
<td>FIT</td>
<td>62.6%</td>
<td>94.3%</td>
</tr>
<tr>
<td>Advanced neoplasms</td>
<td>MonoHaem</td>
<td>FIT</td>
<td>25.6%</td>
<td>97.7%</td>
</tr>
<tr>
<td>Advanced neoplasms</td>
<td>Immudia HemSp</td>
<td>FIT</td>
<td>97.7%</td>
<td>98.8%</td>
</tr>
</tbody>
</table>

Table 12: Summary table from the study by Burch et al for direct comparisons for advanced neoplasms

<table>
<thead>
<tr>
<th>Sensitive GT vs FIT</th>
<th>Sensitivity vs Sensitivity</th>
<th>Specificity vs Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOS,Flexsure</td>
<td>63.3%,79.1%</td>
<td>90.1%,96.9%</td>
</tr>
<tr>
<td>HOS,Immudia HemeSp</td>
<td>78.6%,68.2%</td>
<td>86.7%,94.4%</td>
</tr>
</tbody>
</table>

Table 13: Summary table from the abstracted results of the study by the Australian MSAC for the diagnosis of advanced neoplasms

<table>
<thead>
<tr>
<th>Condition</th>
<th>FIT vs Sensitive or Standard GT</th>
<th>Sensitivity vs Sensitivity</th>
<th>Specificity vs Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced neoplasms</td>
<td>Hemeselect vs Hemoccult (Standard GT)</td>
<td>77.1%,30%</td>
<td>93.5%,97.8%</td>
</tr>
<tr>
<td>Advanced neoplasms</td>
<td>FecatwinSensitive/FecaWIA vs Hemoccult (Standard GT)</td>
<td>50%,83.3%</td>
<td>92.2%,97.1%</td>
</tr>
<tr>
<td>Advanced neoplasms</td>
<td>Hemeselect vs Hemoccult Sensa (Sensitive GT)</td>
<td>79.4%,68.8%</td>
<td>94.4%,86.7%</td>
</tr>
</tbody>
</table>

Evidence from randomized controlled trials

HSAC identified one large RCT conducted on an average risk population in the Netherlands (N = 20623) by van Rossum and colleagues. In this study, 6197 individuals tested with OC-Sensor (FIT) were compared with 4638 individuals who
were tested with Hemoccult II (standard GT). Colonoscopy was used as the criterion standard for the diagnosis of advanced neoplasms. Colonoscopy was only conducted on positive individuals. Thus the sensitivity could not be calculated, but the authors reported specificity, positive predictive value (PPV), positivity and detection rate for advanced neoplasms. The authors concluded that standard GT as a screening tool for average risk adults for CRC was likely to lower estimation of the prevalence of CRC (van Rossum, et al., 2008). See Table 14 for more information.
## Table 14: Summary of the key findings from the van Rossum Study

<table>
<thead>
<tr>
<th>Author &amp; year</th>
<th>Study type and research question</th>
<th>Methods</th>
<th>Key findings</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>van Rossum, 2008</td>
<td>Randomized controlled trial between GT and FIT to identify whether FIT was superior to GT in an average risk elderly population between 50-75 years age for the following outcomes: participation rate, positivity, detection rate, specificity, and PPV.</td>
<td>This was a population based randomized controlled trial where randomization was done at the community level. The target population was individuals between 50-75 years of age, both genders in the Netherlands. The participants were provided with testing kits with instructions. The two screening procedures tested in this study were GT (Hemoccult-II) and FIT (OC-Sensor). The reference standard used in this study was colonoscopy under either midazolam or propofol anaesthesia. Colonoscopies were only conducted on those individuals who tested positive for either test. The outcome variables compared in this study included specificity of the two tests, PPV, Detection Rate, and Positivity.</td>
<td>A total of 20623 individuals participated in the trial (10,301 in the GT arm and 10,322 in the FIT arm). Rate of participation was 46.9% for GT and 59.6 for FIT; the difference was 12.7% (95% CI: 11.4%-14.1%). FIT had lower specificity (97.8% vs 99.1% for GT) and lower PPV (51.8% vs 55.3% for GT). FIT also had higher Positivity (4.54% vs 2.12% for GT) in this population. Detection rate for cancer and advanced polyps were higher in FIT compared to GT (2.35% vs 1.17% in GT). The comparisons were made for advanced neoplasms.</td>
<td>The authors concluded that compared to FIT, GT was more likely to lead to lower estimation of the prevalence of colorectal cancer in a screening population because of lower detection rates. However, significantly less specificity and lower PPV were also likely to lead to higher positivity rates when FIT was used as a screening tool, leading to larger numbers of colonoscopies. However, they were less likely to be offset with more case finding as the PPV for the test was lower as well. In the absence of sensitivity data, a definite choice was not clear based on this RCT.</td>
</tr>
</tbody>
</table>
Evidence from observational epidemiological studies

HSAC identified three observational epidemiological studies that compared different FIT with different types of GT. In a large prospective study, Allison and colleagues (2007) studied 5841 average risk individuals in the United States to evaluate the performance of Flexsure OBT (FIT) with that of Hemoccult Sensa (sensitive GT) as well as the performance of both tests taken together. For advanced neoplasms, they found that Hemoccult Sensa (sensitive GT) had higher sensitivity at 43.1% (95% CI: 34.7%-51.8%) than Flexsure (FIT) with sensitivity figures of 33.1% (95% CI: 24.9%-42.3%). However, Hemoccult Sensa was less specific for these groups of lesions at 90.7% (95% CI: 89.9%-91.5%) compared to Flexsure at 97.5% (95% CI: 97%-97.9%). The point to note is that Flexsure (FIT) had significantly higher PPV 23.1% (17.2%-30.3%) than HOS at 10.1% (7.1%-12.9%). This study supported FIT performing better than sensitive GT (Allison, et al., 2007).

Alicia Smith and colleagues (2006) conducted pairwise comparison of positivity rates between Insure (FIT) and Hemoccult Sensa (sensitive GT) in 2351 screening cohorts and 161 symptomatic individuals (diagnostic cohorts) in Australia. For the screening cohort alone, they found FIT had higher sensitivity but lower specificity than GT for advanced neoplasms. They commented that FIT was significantly better at detecting advanced neoplasms; the true positive rate (TPR) for advanced neoplasms was higher with FIT than sensitive GT (44.4% vs. 24.2%), and PPV for FIT and sensitive GT for advanced neoplasms were 26% and 20.2%, respectively. They concluded that the brush-sampling FIT (Insure) was more sensitive for advanced neoplasms than a sensitive GT (Smith, et al., 2006).

In Israel, Rozen and colleagues (2008) compared OC-Micro (FIT) with Hemoccult II Sensa (sensitive GT). In this analysis, for OC-Micro, they varied the threshold of detection in steps of 25 ng/ml, stepping up the threshold values of FIT from low to high. For OC-Micro threshold set at detecting haemoglobin at levels of 50 mg/100 g faeces, they found that FIT had a sensitivity of 75% while sensitive GT had sensitivity of 53.1%; FIT had specificity of 86.2% and sensitive GT had specificity of 59.4% for advanced neoplasms. Compared to a PPV of 12.1% for sensitive GT for advanced neoplasms in this study, the PPV for FIT was 38.1%, leading to a difference in the number of colonoscopies needed to confirm one case of advanced neoplasm. Using HOS about eight colonoscopies would be required in positive patients to confirm one case of advanced neoplasm while using OC-Micro would required about two colonoscopies (P Rozen, et al., 2008). In summary, all three observational epidemiological studies found that one brand or other of FIT had higher PPV than the Hemoccult Sensa brand of GT. The results are tabulated in Tables 15-17.
Comparison of diagnostic accuracy between immunochemical and guaiac based faecal occult blood tests for colorectal cancer detection

Table 15: Abstracted data from the study by van Rossum et al for the diagnosis of advanced neoplasms

<table>
<thead>
<tr>
<th>Test</th>
<th>Type</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NND</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hemoccult II</td>
<td>Standard GT</td>
<td>99% (98.8%-99.3%)</td>
<td>77.7% (69.6%-85.7%)</td>
<td>1.29</td>
<td></td>
</tr>
<tr>
<td>OC-Sensor</td>
<td>FIT</td>
<td>97.8% (97.4%-98.1%)</td>
<td>77.9% (73.0%-82.7%)</td>
<td>1.28</td>
<td></td>
</tr>
</tbody>
</table>

Table 16: Abstracted data from the study by Allison et al

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Test</th>
<th>Type</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced neoplasms</td>
<td>HOS</td>
<td>Sensitive GT</td>
<td>43.1% (34.7%-51.8%)</td>
<td>90.7% (89.9-91.5%)</td>
</tr>
<tr>
<td>Advanced neoplasms</td>
<td>Flexsure OBT</td>
<td>FIT</td>
<td>33.1% (24.9%-42.3%)</td>
<td>97.5% (97%-97.9%)</td>
</tr>
<tr>
<td>Advanced neoplasms</td>
<td>Results of both tests taken together</td>
<td></td>
<td>26.1% (19.2%-34.4%)</td>
<td>98.5% (98.1%-98.8%)</td>
</tr>
</tbody>
</table>

Table 17: Abstracted data from the study by Smith et al

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Test</th>
<th>Type</th>
<th>Sens</th>
<th>Spec</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced neoplasms</td>
<td>Insure</td>
<td>FIT</td>
<td>65.4%</td>
<td>95.8%</td>
</tr>
<tr>
<td>Advanced neoplasms</td>
<td>Hemoccult II Sensa</td>
<td>Sensitive GT</td>
<td>30.7%</td>
<td>96.3%</td>
</tr>
</tbody>
</table>

Smith et al. (2006) conducted a population-based study of the relative effectiveness of immunochemical FOBT (Insure) versus guaiac based FOBT (Hemoccult II Sensa) in an Australian context. This was a cross-sectional study where people in urban Southern Adelaide were selected as follows:

(a) A self-selected community of members who were asked to volunteer for testing with both immunochemical and guaiac based FOBTs. Within this group, those who tested positive with either test were then requested for follow up with colonoscopy.

(b) The other group was based on those who attended a clinic in a nearby hospital because they perceived that they were at high risk of colonic cancer. In this group, those who had history of other cancer or family history of CRC were offered colonoscopy irrespective of their status of FOBT test.

(c) The third group of individuals were those who had symptoms suggestive of CRC, and they were drawn from the community as well as from the hospital attendees. In this analysis, however, only results from those individuals who were included in the screening programme were analysed.

Altogether 2351 individuals were subjected to the screening programme. Of these 431 individuals (18.3%) underwent colonoscopies. This analysis is based on a subsample of 225 colonoscopy cases. In this study, when results of the screening population alone were considered, the positivity of FIT was 5.57 while that of GT was 3.99.
When Significant Adenoma and CRC were considered together, the sensitivity of FIT was higher than that of sensitive GT (65.4% vs 30.6%). For these groups of lesions, PPV of FIT was higher than that of sensitive GT (25.9% vs 20.2%), and consequently likelihood ratio was higher as well (15.4 vs 8.28). The specificity of GT was higher than that of FIT (96.3% vs 95.7%) (Table 17).

For the diagnosis of cancer, sensitivity of FIT was higher than that of GT (82.4% vs 47.1%), LR was higher for FIT (16.4 vs 12.8), and Detection Rate was higher for FIT (0.59% vs 0.34%)

Rozen and colleagues (2008) reported results of a study of head to head comparison of FIT with GT. In this study, Rozen et al. (2008) compared FIT (OC-Micro) with GT (Hemoccult Sensa) among at-risk individuals. A total of 330 above- or at-average risk individuals were offered Hemoccult Sensa (GT) and OC-Micro (FIT) tests and were followed up over a period of three months using colonoscopy and biopsy as necessary. Additionally, the investigators studied the screening properties of FIT by increasing the threshold levels ranging from 50 ng/ml through to 200 ng/ml in steps of 50 ng/ml. They found GT identified 138 positive Significant Adenomas and CRCs of which 17 were confirmed by colonoscopy. GT also identified 192 negative patients, 15 of whom were found to be positive by colonoscopy. In this high risk and symptomatic group of individuals, use of GT resulted in a net positivity of 41.8%, and sensitivity of about 53.1%, specificity of 59.4%, and positive predictive value of 12.3%, leading to about eight colonoscopies required to identify one person with CRC or Significant Adenoma.

On the other hand, at 50 ng/ml threshold level, sensitivity was around 75%, specificity was 86.2%, and overall positivity was 19.6%. Based on the PPV values, about two colonoscopies would be required to identify one person with Significant Adenoma or CRC. Moreover, when the two tests were conducted together, this led to marginally increased sensitivity and specificity over one FIT test and a comparable number of colonoscopies to identify one true positive out of the total number of positive individuals identified by the combination screening test (Table 18).

**Table 18: Abstracted data showing effect of varying the threshold for FIT values, in this case OC-Micro, from the study by Rozen et al.**

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Test</th>
<th>Threshold</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NND</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced neoplasms</td>
<td>HOS (Hemoccult Sensa, Sensitive GT)</td>
<td>50</td>
<td>53.1</td>
<td>59.4</td>
<td>12.3</td>
<td>8.1</td>
</tr>
<tr>
<td>Advanced neoplasms</td>
<td>OC Micro (FIT)</td>
<td>75</td>
<td>68.8</td>
<td>90.6</td>
<td>52.8</td>
<td>1.9</td>
</tr>
<tr>
<td>Advanced neoplasms</td>
<td>OC Micro (FIT)</td>
<td>100</td>
<td>68.8</td>
<td>93.3</td>
<td>60</td>
<td>1.7</td>
</tr>
<tr>
<td>Advanced neoplasms</td>
<td>OC Micro (FIT)</td>
<td>125</td>
<td>56.3</td>
<td>95</td>
<td>64</td>
<td>1.6</td>
</tr>
<tr>
<td>Advanced neoplasms</td>
<td>OC Micro (FIT)</td>
<td>150</td>
<td>56.3</td>
<td>95.3</td>
<td>64</td>
<td>1.6</td>
</tr>
<tr>
<td>Advanced neoplasms</td>
<td>OC Micro (FIT)</td>
<td>200</td>
<td>53.1</td>
<td>96</td>
<td>65.2</td>
<td>1.5</td>
</tr>
</tbody>
</table>
Results from simple pooled data analysis

Data from the primary studies were abstracted and used to compare the performances of FIT versus standard or sensitive GT. In pooling these data together, the following measures of performances of screening tests were taken into account: sensitivity, specificity, Positive Predictive Value, and Numbers Needed to Detect. For each of these measures median, 25th percentile and 75th percentile figures were reported. For the group of studies pooled under standard or sensitive GT, the median sensitivity was 50% (25th Percentile values: 43.1%, 53.1%), the median specificity was 90.7% (80.1%, 97.1%), the median PPV was 20.2% (12.7%, 55.2%) and the corresponding median NND was 4.95 (1.81, 7.86) For all studies pooled under FIT, the median sensitivity was 67.5% (42.6%, 73.7%), the median specificity was 91.1% (80.1%, 97.7%), the PPV was 38.1% (28.2%, 54.1%), and the corresponding NND was 2.62 (1.86, 3.53) (Tables 19-21).
### Table 19: Summary of key studies and key results

<table>
<thead>
<tr>
<th>Author &amp; year</th>
<th>Study Type and Research Question</th>
<th>Methods</th>
<th>Key findings</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allison, 2007</td>
<td>Relative effectiveness of guaiac based FOBT (Hemoccult Sensa) versus immunochemical FOBT (FIT, Flexsure) in an average risk US adult population; Primary Study based on cross sectional survey of all participants in the study</td>
<td>7394 at-average risk individuals were offered three test cards containing GT (Hemoccult Sensa), FIT (Flexsure OBT), and combination of GT and FIT kits. Of 7394 individuals, 5932 individuals were screened by FOBT. This reference standard for the screening programme was based on a combination of Colonoscopy &amp; Sigmoidoscopy. Colonoscopy was done for individuals with positive results in either GT or FIT. Those who tested negative in both tests were followed by Flexible Sigmoidoscopy. Those who tested positive in all three tests (GT, FIT, and combination) were deemed as Positive in both tests, otherwise they were termed as negative in both tests.</td>
<td>Overall 584 people reported testing positive with GT, and 173 people reported positive with FIT. For CRC and Significant Adenoma, the sensitivity of FIT was 33.1% and that of GT was 43.1%; However, the PPV of FIT was higher in this study than GT.</td>
<td>This study identified that for colorectal cancer, based on likelihood ratio estimates, FIT (Flexsure) was a better screening tool than GT (Hemoccult Sensa). However, for small adenomas and adenomas &gt; 1 cm in size, GT as a screening tool performed better.</td>
</tr>
<tr>
<td>Author &amp; year</td>
<td>Study Type and Research Question</td>
<td>Methods</td>
<td>Key findings</td>
<td>Conclusion</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------------------</td>
<td>---------</td>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>Smith, 2006</td>
<td>This was a population based study of the relative effectiveness of immunochemical FOBT (Insure) versus guaiac based FOBT (Hemoccult II Sensa) in an Australian population</td>
<td>This was a population based cross sectional study where people in urban Southern Adelaide were divided into two groups: (a) a self selected community group who were asked to volunteer for testing with both immunochemical and guaiac based FOBTs. Within this group, those who tested positive with either test were then requested for follow up with colonoscopy. The other group was based on those who attended a clinic in a nearby hospital because they perceived that they were at high risk of colonic cancer. In this group, those who had history of other cancer or those with family histories of CRC were offered colonoscopy irrespective of their status of FOBT test. The third group of individuals were those who had symptoms suggestive of CRC, and they were drawn both from the community as well as from the hospital attendee. In this analysis, however, only results from those individuals who were included in the screening programme were analysed.</td>
<td>Altogether 2351 individuals were subjected to the screening programme. Of these 431 individuals underwent colonoscopies. This analysis is based on a subsample of 225 colonoscopy cases. In this study, when results of screening population alone were concerned, the positivity of FIT was 5.57 while that of GT was 3.99. However, for combination of CRC and Significant Adenoma (adenoma that were 10 mm or more in size), the sensitivity of FIT was higher than that of GT (65.4% vs 30.6%), PPV of FIT was higher than that of GT (25.9% vs 20.2%), and consequently likelihood ratio was higher as well (15.4 vs 8.28). The specificity of GT was higher than that of FIT (96.3% vs 95.7%). For the diagnosis of cancer, sensitivity of FIT was higher than that of GT (82.4% vs 47.1%), LR was higher for FIT (16.4 vs 12.8), and Detection Rate was higher for FIT (0.59% vs 0.34%).</td>
<td>Based on these parameters, FIT performed better than GT in the screening of both significant colorectal adenomas and colorectal cancers.</td>
</tr>
</tbody>
</table>
## Table 19: Summary of key studies and key results (continued)

<table>
<thead>
<tr>
<th>Author &amp; year</th>
<th>Study Type and Research Question</th>
<th>Methods</th>
<th>Key findings</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rozen, 2008</td>
<td>This was a head to head comparison between FIT (OC-Micro) and GT (Hemoccult Senza) among at risk individuals</td>
<td>The investigators conducted a head to head comparison between GT (Hemoccult Senza) and FIT (OC-Micro) in Israel. A total of 330 above-average risk individuals were offered Hemoccult Senza (GT) and OC-Micro (FIT) tests and were followed up over a period of three months using colonoscopy and biopsy as necessary. The investigators provided data to calculate sensitivity, specificity, PPV, LR, for comparing between FIT and GT tests. The investigators also studied the screening properties of FIT at threshold levels ranging from 50 ng/ml through to 200 ng/ml</td>
<td>GT identified 138 positive Significant Adenomas and CRCs of which 17 were confirmed by colonoscopy. GT also identified 192 negative patients, 15 of whom were found to be positive by colonoscopy. This resulted in a net positivity of 41.8%, and sensitivity of about 53.1%, specificity of 59.4%, and positive predictive value of 12.3%, leading to about 8 colonoscopies required to identify one person with cancer or Significant Adenoma among all positive patients identified by the GT. On the other hand, at 50 ng/ML threshold level, the sensitivity was around 75%, specificity was 86.2%, and overall positivity was 19.6%, and based on the PPV values, about 2 colonoscopies would be required to identify one person with Significant Adenoma or cancer based on FIT. However, when the two tests were conducted together this led to marginally increased sensitivity and specificity over one FIT test, and a comparable number of colonoscopies to identify one true positive out of the total number of positive individuals identified by the combination screening test.</td>
<td>FIT was found to have higher specificity although had lower sensitivity and lead to fewer colonoscopies. They concluded that compared to GT, FIT had better profiles in terms of sensitivity and specificity; additionally, it had also resulted in fewer number of colonoscopies required to confirm one positive case (the reduction was almost 75% reduction in the number of colonoscopies in their trial).</td>
</tr>
</tbody>
</table>
**Table 20: Comparison between GT and FIT for significant adenoma and CRC for sensitivity, specificity, and NND**

<table>
<thead>
<tr>
<th></th>
<th>Median</th>
<th>25th percentile</th>
<th>75th percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FIT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>67.5</td>
<td>42.6</td>
<td>73.6</td>
</tr>
<tr>
<td>Specificity</td>
<td>91.1</td>
<td>80.1</td>
<td>97.1</td>
</tr>
<tr>
<td>NND</td>
<td>2.62</td>
<td>1.85</td>
<td>3.53</td>
</tr>
<tr>
<td><strong>Standard or Sensitive GT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>50.1</td>
<td>43.1</td>
<td>53.1</td>
</tr>
<tr>
<td>Specificity</td>
<td>90.7</td>
<td>76.2</td>
<td>95.3</td>
</tr>
<tr>
<td>NND</td>
<td>4.9</td>
<td>1.81</td>
<td>7.86</td>
</tr>
</tbody>
</table>
### Table 21: Pooled data from the primary studies

<table>
<thead>
<tr>
<th>Author, Year</th>
<th>Brand Name Method)</th>
<th>Cut-off for FIT</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>TP</th>
<th>FP</th>
<th>TN</th>
<th>FN</th>
<th>FNR</th>
<th>TPR</th>
<th>PPV</th>
<th>NND</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allison, 2007</td>
<td>Flexsure OBT (FIT)</td>
<td>33.1</td>
<td>97.5</td>
<td>40</td>
<td>133</td>
<td>81</td>
<td></td>
<td></td>
<td>0.378505</td>
<td>0.330579</td>
<td>23.1</td>
<td>4.33</td>
</tr>
<tr>
<td>Allison, 2007</td>
<td>Hemoccult Sensa (Sensitive GT)</td>
<td>43.1</td>
<td>90.7</td>
<td>59</td>
<td>525</td>
<td>78</td>
<td></td>
<td></td>
<td>0.129353</td>
<td>0.430657</td>
<td>10.1</td>
<td>9.89</td>
</tr>
<tr>
<td>Smith, 2006</td>
<td>Insure (FIT)</td>
<td>Not stated</td>
<td>65.4</td>
<td>95.8</td>
<td>34</td>
<td>97</td>
<td>2202</td>
<td>18</td>
<td>0.156522</td>
<td>0.653846</td>
<td>25.9</td>
<td>3.86</td>
</tr>
<tr>
<td>Smith, 2006</td>
<td>Hemoccult II (Standard GT)</td>
<td>Not Stated</td>
<td>30.7</td>
<td>96.3</td>
<td>19</td>
<td>75</td>
<td>2214</td>
<td>43</td>
<td>0.364407</td>
<td>0.306452</td>
<td>20.2</td>
<td>4.95</td>
</tr>
<tr>
<td>van Rossum, 2008</td>
<td>Hemoccult II (Standard GT)</td>
<td>xx</td>
<td>na</td>
<td>57</td>
<td>46</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>55.4</td>
<td>1.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>van Rossum, 2006</td>
<td>OC-Sensor (FIT)</td>
<td>100 ng/ml of faeces</td>
<td>na</td>
<td>57</td>
<td>46</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>55.4</td>
<td>1.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rozen, 2008</td>
<td>Hemoccult Sensa (Sensitive GT)</td>
<td>53.1</td>
<td>59.4</td>
<td>17</td>
<td>121</td>
<td>177</td>
<td>15</td>
<td></td>
<td>0.110294</td>
<td>0.53125</td>
<td>12.3</td>
<td>8.1</td>
</tr>
<tr>
<td>Rozen, 2008</td>
<td>OC-Micro (FIT)</td>
<td>50 ng/ml</td>
<td>75</td>
<td>86.2</td>
<td>24</td>
<td>41</td>
<td>257</td>
<td>8</td>
<td>0.163265</td>
<td>0.75</td>
<td>38.1</td>
<td>2.62</td>
</tr>
<tr>
<td>Rozen, 2008</td>
<td>OC-Micro (FIT)</td>
<td>100 ng/ml</td>
<td>68.7</td>
<td>96.3</td>
<td>22</td>
<td>10</td>
<td>288</td>
<td>10</td>
<td>0.3125</td>
<td>0.687</td>
<td>68.7</td>
<td>1.45</td>
</tr>
</tbody>
</table>

Comparison of diagnostic accuracy between immunochemical and guaiac based faecal occult blood tests for colorectal cancer detection
Comparison of diagnostic accuracy between immunochemical and guaiac based faecal occult blood tests for colorectal cancer detection
Summary and Conclusions

Summary of evidence for evidence review
To summarize the main findings of this review – (1) systematic reviews of direct comparison between different brands of FIT and GT did not find sufficient evidence in favour of one or other brand of FIT or FIT as a class of screening tools over different brands of GT or GT as a class of screening tool; (2) one RCT found similar figures for sensitivity, specificity and PPV between OC-Sensor (FIT) and Hemoccult Sensa (GT), and (3) Observational Epidemiological Studies with three different brands of FIT (Flexsure, Insure, and OC-Media) found that while FIT had lower sensitivity compared to GT (Hemoccult Sensa or Hemoccult II Sensa), all three studies found higher PPV compared to GT; (4) When the data of all the tests from the different studies were pooled together, FIT as a class of screening test had higher median sensitivity, specificity, PPV, and therefore lower numbers needed to detect (NND) compared to GT.

Discussion and conclusions
These findings suggest that although use of FIT might lead to a higher percentage of individuals being identified as positive for CRC/Significant Adenoma, it might result in more optimized use of colonoscopy as a criterion test.

The purpose of this review was to evaluate the relative screening properties of guaiac versus immunochemical FOBTs, in order to inform a process of selection of the better of the two tools for screening of CRC. While the systematic reviews and RCTs were inconclusive whether FIT or GT were likely to be better alternatives, observational epidemiological studies found that even with low sensitivity, all three brands of FIT (Flexsure, Insure, and OC-Micro) were superior in PPV and NND when compared to Hemoccult Sensa or Hemoccult II Sensa. This indicated that performance wise, as a single test alternative, FIT might be deemed to make better use of colonoscopy resources, even though they may result in higher overall Positivity Rates.

However, the results of this review need to be interpreted in the light of several limitations and restrictions of the approach taken. Firstly, this review used sensitivity, specificity, positive predictive value, and number of colonoscopies needed to confirm one case of advanced neoplasm (NND or numbers needed to detect) as parameters to decide if one test was superior to another. While these measures taken together provide some guidance in deciding one test is superior to another, they are by no means prescriptive. Sensitivity and specificity of screening tests are measures of their accuracies of ruling in and ruling out of the presence of the target disease respectively, and therefore their appropriateness as choices of screening tests. Two other measures – positivity and PPV/NNDs - might be important in informing the decision of how useful a screening test is in the real world setting, or help in the selection of one or another test. Positivity refers to the proportion of individuals identified as positive out of all people tested with the screening programme, and a high positivity (as seen with tests that are highly sensitive) indicate large numbers of individuals to be followed up with the confirmatory test. If positivity is also associated with a high positive predictive value, this indicates high True Positive
Rates. From an operational perspective, implication of a high PPV is that few colonoscopies will be required to identify one true case of advanced adenoma for a particular screening option, indicating more efficient use of resources (numbers needed to detect). Therefore, even if a screening test has high Positivity Rate due to its high sensitivity figures, if the corresponding figures of PPV are also high (and corresponding NND are low), that essentially indicates low False Positive Rate and therefore better use of resources used for confirmation of the screening diagnoses.

Secondly, other attributes of a screening option that can potentially play important roles in deciding its adoption were not studied nor reported in this review. These include relative costs of the screening options, uptake rates or adoption by people to whom the screening options are offered, and in the case of CRC, issues around preparation of the screening population and interpretation of the results. These are important issues that need to be taken into account for making a decision of adopting one or other screening option.

Thirdly, this review has attempted to evaluate two rival screening options (guaiac based tests versus faecal immunochemical tests) in terms of their screening performances only. It did not identify studies that had compared whether sensitive guaiac was superior to standard guaiac, nor did it identify studies that had compared different types of FIT tests. Only one study in this review had data about comparison of different threshold levels of FIT (OC-Micro) with each other and with GT (Hemoccult II Sensa). In this study by Rozen and colleagues in Israel, OC-Micro was found to have lower test sensitivity at levels 125 mg/100 g of faeces (equivalent of ng/ml), but even at these thresholds, the specificity of OC-Micro (FIT) was better than that of Hemoccult II Sensa, as was the corresponding PPV, indicating that even when the threshold of detection is set at higher levels than usual, OC-Micro could still be better in terms of specificity, PPV and NND. Therefore, while this review found some evidence in support of superior performance of FIT tests as a family of screening options over sensitive guaiac tests (and some indirect evidence of superiority of sensitive guaiac tests over standard guaiac tests because of their designed increased sensitivity to identify pseudoperoxidase activity in blood at lower concentrations), the results of the study cannot provide any indication whether one form of FIT is superior to another. Other, newer forms of FOBT and other types of screening tests were not considered as part of this review.

Finally, the time frame for this review was 2004 - 2008. This time frame was selected because comprehensive systematic reviews based on research on CRC screening prior to 2004 were already available and were included in this review. In conclusion, this review found some support from observational epidemiological studies that different forms of FIT (Flexsure, OC-Micro and Insure) were superior to GT for screening of CRC.
Comparison of diagnostic accuracy between immunochemical and guaiac based faecal occult blood tests for colorectal cancer detection

References


Comparison of diagnostic accuracy between immunochemical and guaiac based faecal occult blood tests for colorectal cancer detection
Comparison of diagnostic accuracy between immunochemical and guaiac based faecal occult blood tests for colorectal cancer detection
Glossary

**Absolute risk reduction (ARR)**   The difference between adverse outcomes in the treated and placebo groups in a clinical trial.

**Absolute risk**   The observed or the calculated risk of an event in a population.

**Acute services**   Services for urgent conditions that need immediate treatment.

**Age adjusted rates**   Mortality or morbidity rates in which there has been adjustment for differences in age distribution of the populations being compared.

**Age standardisation**   A procedure for adjusting rates designed to minimise the effects of differences in age composition when comparing rates for different populations.

**Ambulatory care**   Health care provided in other than an inpatient setting.

**Analysis of variance (ANOVA)**   A statistical analysis involving the comparison of variance reflecting different sources of variability.

**Before and after study**   A situation in which the investigator compares outcomes before and after the introduction of a new intervention.

**Benchmarking**   The process of comparing the prices, quality or scope of services against those of other similar services or against common reference points or standards.

**Bias**   Deviation of results or inferences from the truth, or processes leading to such deviation. Any trend in the collection, analysis, interpretation, publication, or review of data that can lead to conclusions that are systematically different from the truth.

**Blinded study**   A study in which observers and/or subjects are kept ignorant of the group to which they are assigned. When both observers and subjects are kept ignorant, the study is referred to as double blind.

**Budget holding**   A system where managers have a fixed budget for a defined population and must meet the costs of an agreed set of services used by that population.

**Capitation**   A system of paying providers a defined price for each consumer who is registered with a provider.

**Case control study**   An epidemiological study involving the observation of cases (persons with the disease, such as cervical cancer) and a suitable control (comparison, reference) group of persons without the disease. The relationship of an attribute to the disease is examined by comparing retrospectively the past history of the people in the two groups with regard to how frequently the attribute is present. See also nested case control.
**Case fatality rate**  The proportion of cases of a specified condition which are fatal within a specified time.

**Case management**  All health professionals’ activities to ensure coordination of services for the patient.

**Case series**  A descriptive study of a subset of a defined population (i.e. a single patient or group of patients) which aims to describe the association between factors or attributes which the sample are exposed to, and the probability of occurrence of a given disease or other outcome. Case series are collections of individual case reports, which may occur within a fairly short period of time.

**Casemix**  The distribution and different types of patients cared for in a health care facility.

**Chalmers scale**  A validated scale that examines the methodological quality of epidemiologic research. The scale uses a scoring system that weights particular aspects of study design and reporting. Greatest emphasis is placed on blinding and other variables considered including the analytic techniques used, control of bias, testing procedures and presentation of results.

**Cohen classification**  A system that categorises the pooled effect size into small, medium and large categories. A small effect accounts for less than or equal to 1% of the variance of the population, a medium effect accounts for 1% to less than or equal to 5.9% of the population variance, and a large effect accounts for between 5.9% and 13.8% of the variance.

**Cohort study**  The analytic method of epidemiologic study in which subsets of a defined population can be identified who are, have been, or in the future may be exposed or not exposed in different degrees, to a factor or factors hypothesised to influence the probability of occurrence of a given disease or other outcome. Studies usually involve the observation of a large population, for a prolonged period (years), or both.

**Community care**  Corresponds to ambulatory and domiciliary services provided other than through hospitals to patients who are resident at their home, hotel, prison, barrack etc.

**Confidence interval**  The computed interval with a given probability, e.g. 95%, that the true value of a variable such as a mean, proportion, or rate is contained within the interval. The 95% CI is the range of values in which it is 95% certain that the true value lies for the whole population.

**Confounder**  A third variable that indirectly distorts the relationship between two other variables, because it is independently associated with each of the variables.

**Confounding**  A situation in which the measure of the effect of an exposure on risk is distorted because of the association of exposure with other factor(s) that influence the outcome under study.
**Cost benefit analysis**  An economic analysis in which the costs of medical care and the benefits of reduced loss of net earnings due to preventing premature death or disability are considered.

**Cost effectiveness (CE)**  Involves the relationship between costs and effects, providing information on whether a technology is being delivered to those who would benefit from it with an optimal use of resources. It is expressed as a ratio of the effects (number of lives saved, number of disability days avoided) obtained for a specific cost (expressed in dollars). For example, the numerator may be the difference in lifetime costs between one intervention and another, while the denominator may be the difference in life expectancies associated with the two interventions. Low cost effectiveness ratios are desirable.

**Cost minimisation analysis**  A particular type of cost effectiveness analysis in which it is assumed the outcome is the same in all comparison groups. The focus is, therefore, on the comparative costs of different interventions.

**Coverage**  The number, percent, or proportion of eligible people reached by a programme.

**Critical care pathway**  Schedules in patient care for coordinated treatment.

**Cross-sectional study**  A study that examines the relationship between diseases (or other health related characteristics), and other variables of interest as they exist in a defined population at one particular time.

**Day patient**  A person who is admitted and discharged from hospital on the same day.

**Demand driven services**  Services purchased for an unlimited fee-for-service basis according to demand.

**Descriptive study**  A study concerned with, and designed only to describe the existing distribution of variables, without regard to causal or other hypotheses.

**Diagnostic related group**  A grouping of cases on the basis of similar cost within broader groupings relating to the same or similar organ or system of the body.

**Discounting**  In cost effectiveness studies, future dollar costs and benefit streams are reduced or “discounted” by a percentage to reflect the fact that money spent or saved in the future should not weigh as heavily in programme decisions as dollars spent or saved now.

**Discrepancy study**  A study involving the verification of only discordant diagnoses of two interventions (such as two screening tests), that is, where one test provides a positive result and the other provides a negative result.
**Dominance** In cost effectiveness studies, an alternative is eliminated by dominance if it is both less effective and more costly than (i.e. dominated by) at least one other alternative.

**Ecological study** A study in which the units of analysis are populations or groups of people, rather than individuals.

**Effectiveness** A measure of the extent to which a specific intervention, procedure, regimen, or service, when deployed in the field in routine circumstances, does what it is intended to do for a specified population.

**Efficiency** The effects or end results achieved in relation to the effort expended in terms of money, resources and time. The extent to which the resources used to provide a specific intervention, procedure, regimen, or service of known efficacy and effectiveness are minimised.

**Elective services** Non-urgent services for conditions which do not need immediate treatment. This includes services for patients with semi-urgent or non-life-threatening chronic conditions that tend to be stable or slowly deteriorate over time.

**Evidence based** Based on valid empirical information.

**Extended dominance** In cost effectiveness studies, an alternative is eliminated by extended dominance if it has a higher cost effectiveness ratio than a more effective option.

**Final truth determination** Use of a reference standard to provide an accurate or “truth” diagnosis for verification of positive and negative diagnoses by a screening or diagnostic test (see also “reference standard”).

**Generalisability** Applicability of the results to other populations.

**Grey literature** That which is produced by all levels of government, academics, business and industry, in print and electronic formats, but which is not controlled by commercial publishers.

**High risk groups** Usually refers to groups of women that have been identified as having a higher than expected, or higher than average for the population as a whole, incidence of the disease in question.

**Histology** The microscopic study of the minute structure and composition of tissues.
Hospitalisations A term used as an indicator of morbidity of diseases in a community. A hospitalisation in New Zealand health statistics includes inpatients who leave hospital to return home, transfer to another hospital or institution, or die in hospital after formal admission. That is a count of episodes of care rather than individuals.

Incidence The number of new events (cases; e.g. of disease) occurring during a certain period, in a specified population.

Indicator An item of quantitative or qualitative information reported to enable the monitoring of a condition or the performance of an organisation.

Inpatient A person admitted to hospital for medical, surgical or psychiatric treatment, observation or care, which spends at least one night in the hospital. A healthy person accompanying a sick person is included if formally admitted as a boarder.

Intangible costs The costs of suffering to the patients and their carers.

Integrated care A system of health care which aims to integrate the delivery of health care for a defined population, with explicit contracts and protocols developed by one overall provider organisation.

Intention to treat A method for data analysis in a randomised controlled trial in which individual outcomes are analysed according to the group to which they were randomised even if they never received the treatment to which they were assigned.

Intention to treat analysis A method for data analysis in a randomised controlled trial in which individual outcomes are analysed according to the group to which they were randomised even if they never received the treatment to which they were assigned.

Managed care organisation An organisation or service provider which is given responsibility for ensuring that a defined population receives a defined set of services in a co-ordinated way.

Matching The process of making a study group and a comparison group comparable with respect to extraneous factors.

Mean Calculated by adding all the individual values in the group and dividing by the number of values in the group.

Median Any value that divides the probability distribution of a random variable in half. For a finite population or sample the median is the middle value of an odd number of values (arranged in ascending order) or any value between the two middle values of an even number of values.

Meta-analysis The process of using statistical methods to combine the results of different studies. The systematic and organised evaluation of a problem, using information from a number of independent studies of the problem.
Misclassification  The erroneous classification of an individual, a value, or an attribute into a category other than that to which it should be assigned.

Morbidity  Illness.

Mortality  The number of deaths from a specified disease which are diagnosed or reported during a defined period of time in a given population.

Multiple regression  Any analysis of data that takes into account a number of variables simultaneously.

National Minimum Dataset  Data on specific procedures required from service providers by the New Zealand Ministry of Health.

Natural history  The course of a disease from onset to resolution.

Negative predictive value  The probability a person does not have the disease when the screening test is negative.

Nested case control study  A case control study in which cases and controls are drawn from the population in a cohort study. That is, the case control study is “nested” within the cohort study design so that the effects of some potential confounding variables are reduced or eliminated. A case control study can also be nested into a case series study. See also case control study, cohort study, and case series study.

Number needed to treat (NNT)  The number of patients who need to be treated to achieve one additional favourable outcome. Calculated as 1/ARR. If the intervention harmed people, the term would be the number needed to harm.

Odds ratio (OR)  A measure of the degree or strength of an association. In a case control or a cross sectional study, it is measured as the ratio of the odds of exposure (or disease) among the cases to that among the controls.

OECD  Organisation for Economic Co-operation and Development. There are 24 countries in the OECD.

Outpatient  A person who goes to a health care facility for a consultation, and who leaves the facility within three hours of the start of the consultation. An outpatient is not formally admitted to the facility.

Parenteral  Administering a medication by injection through a route other than by alimentary canal (e.g. intramuscularly or intravenously).

Period prevalence  The total number of persons known to have the disease or attribute at any time during a specified period.

Polypharmacy  The administration of two or more drugs together.

Positive predictive value  The probability a person actually has the disease when the screening test is positive.
**Power** The ability of a study to demonstrate an association if one exists.

**Prevalence** The number of events in a given population at a designated time (point prevalence) or during a specified period (period prevalence).

**Primary care** First contact, continuous, comprehensive and coordinated care provided to individuals and populations undifferentiated by age, gender, disease or organ system.

**Providers** Organisations and health professionals providing health services.

**Random sample** A sample that is arrived at by selecting sample units such that each possible unit has a fixed and determinate probability of selection.

**Randomised controlled trial** An epidemiologic experiment in which subjects in a population are randomly allocated into groups to receive or not receive an experimental preventive or therapeutic procedure, manoeuvre, or intervention. Randomised controlled trials are generally regarded as the most scientifically rigorous method of hypothesis testing available in epidemiology.

**Recall bias** Systematic bias due to differences in accuracy or completeness of recall or memory of past events or experiences.

**Reference standard** An independently applied test that is compared to a screening or diagnostic test being evaluated in order to verify the latter’s accuracy. A reference standard, therefore, provides an accurate or “truth” diagnosis for verification of positive and negative diagnoses. It is sometimes described as providing “final truth determination”.

**Relative risk (RR)** The ratio of the risk of disease or death among the exposed to the risk among the unexposed. It is a measure of the strength or degree of association applicable to cohort studies and RCTs.

**Relative risk reduction (RRR)** The proportional reduction in rates of bad events between experimental and control participants in a trial. If there was an increase in the rate of bad events, the term would then be relative risk increase.

**Risk factor** An exposure or aspect of personal behaviour or lifestyle, which on the basis of epidemiologic evidence is associated with a health-related condition.

**Secondary care** Surgical and medical services that are generally provided in a hospital setting. In many cases, access to these services is by referral from a primary care health professional such as a general practitioner.

**Segi’s world population** A standard population which is merely an arbitrary set of figures against which other populations can be standardised to produce comparable rates.

**Selection bias** Error due to systematic differences in characteristics between those who are selected for inclusion in a study and those who are not (or between those compared within a study and those who are not).
**Sensitivity analysis**  A method to determine the robustness of an assessment by examining the extent to which results are affected by changes in methods, values of variables, or assumptions.

**Sensitivity**  Sensitivity is the proportion of truly diseased persons in a screened population who are identified as diseased by a screening test. Sensitivity is a measure of the probability of correctly diagnosing a case, or the probability that any given case will be identified by the test.

**Specificity**  The proportion of truly non-diseased persons who are so identified by a screening test. It is a measure of the probability of correctly identifying a non-diseased person with a screening test.

**Standardised mortality ratio (SMR)**  The ratio of the number of deaths observed in the study group or population to the number that would be expected if the study population had the same specific rates as the standard population, multiplied by 100.

**Systematic review**  Literature review reporting a systematic method to search for, identify and appraise a number of independent studies.

**Utilisation review**  Systematic review of particular procedures to ensure the right thing was done to the right person, in the right places, at the right time and in the right way.

**Variance**  A measure of the variation shown by a set of observations, defined by the sum of the squares of deviation from the mean, divided by the number of degrees of freedom in the set of observations.

**Verification bias**  A type of bias that occurs when a study selectively includes patients for disease verification (or exclusion) by reference standard testing, based on positive or negative results of preliminary testing, or the study test itself. To avoid this, a study should include consecutive patients at risk for a particular disease, and not only a subset who underwent definitive testing.

**Wilson and Jungner’s criteria**  Criteria that should be satisfied in a screening programme.
Appendix A: Included Studies


Comparison of diagnostic accuracy between immunochemical and guaiac based faecal occult blood tests for colorectal cancer detection
Appendix B: Excluded Studies


Comparison of diagnostic accuracy between immunochemical and guaiac based faecal occult blood tests for colorectal cancer detection


Comparison of diagnostic accuracy between immunochemical and guaiac based faecal occult blood tests for colorectal cancer detection


Marcus, A. C., Mason, M., et al. (2005). The efficacy of tailored print materials in promoting colorectal cancer screening: results from a randomized trial involving callers to the National Cancer Institute's Cancer Information Service. 1 (??), 83-104.

Medical Services Advisory. (2004). Faecal occult blood testing for population health screening (Structured abstract). Canberra: Medical Services Advisory Committee (MSAC), 122.


Comparison of diagnostic accuracy between immunochemical and guaiac based faecal occult blood tests for colorectal cancer detection
Appendix C: Quality Checklists for Appraising Interventions

Method of treatment assignment
- Correct, blinded randomisation method described OR randomised, double-blind method stated AND group similarity documented
- Blinding and randomisation stated but method not described OR suspect technique (e.g. allocation by drawing from an envelope)
- Randomisation claimed but not described and investigator not blinded.
- Randomisation not mentioned

Control of selection bias after treatment assignment
- Intention to treat analysis AND full follow-up
- Intention to treat analysis AND <15% loss to follow-up
- Analysis by treatment received only OR no mention of withdrawals
- Analysis by treatment received AND not mention of withdrawals OR more than 15% withdrawals/loss to follow-up/post-randomisation exclusions

Blinding
- Blinding of outcome assessor AND patient and care giver
- Blinding of outcome assessor OR patient and care giver
- Blinding not done

Outcome assessment (if blinding was not possible)
- All patients had standardised assessment
- No standardised assessment OR not mentioned


Comparison of diagnostic accuracy between immunochemical and guaiac based faecal occult blood tests for colorectal cancer detection
Comparison of diagnostic accuracy between immunochemical and guaiac based faecal occult blood tests for colorectal cancer detection
Appendix D: Performances at Various Threshold Levels of Faecal Immunochemical Tests

Comparison of the performances of Hemoccult Sensa at various threshold levels versus Faecal Immunochemical Tests on the basis of efficiency, safety, impact and feasibility

Introduction
This section briefly describes a head to head comparison of a faecal immunochemical test (OC-Micro) with a sensitive guaiac based faecal occult blood test (Hemoccult Sensa). The threshold levels of the test that are used for this comparison are described in a study conducted by Rozen et al (2008) in an above-average risk group of individuals in Israel who underwent head to head comparisons of the performance between the two tests. The purpose of this paper is to examine on the basis of the data presented in the paper to identify the best faecal occult blood test and the best threshold level of detection of advanced neoplasms. Advanced neoplasms are defined as colorectal neoplasms that are either 10 mm or more in diameter, or those neoplasms that show more than 20% villous adenomatous change on histology, irrespective of size. Advanced neoplasm as defined in this paper refers to both Significant Adenoma (i.e. benign tumours that are more than 10 mm in diameter or villous adenoma that might be less than 10 mm in diameter) as well as colorectal cancers.

Methods
The data for the results shown in Table 22 were taken from the study by Rozen and colleagues (2008) on quantitative colonoscopic evaluation of relative efficiencies of Hemoccult sensa (a sensitive GT) and OC-Micro (a faecal immunochemical test). The following measures were used to compare the performances of the two tests:

The following variables were used to compare the relative performances of the two tests at various levels of the faecal immunochemical tests in units of nanograms/ml of detection of hemoglobin in the stool samples – sensitivity of the tests, specificity of the tests, relative sensitivities (Relative Sensitivity = Ratio of sensitivity of sensitive GT over FIT), relative positivities (ratio of positivity of sensitive GT over FIT), 7 changes in the relative sensitivities and positivities. The changes in the number of colonoscopies required were calculated by multiplying the estimated (provisional number of Colonoscopies shown as an illustration) calculated by multiplying the baseline number of colonoscopies (the provisional estimated number of colonoscopies with Hemoccult Sensa for a programme shown here as 4000) with the change in the positivity threshold. In calculating these numbers, performances of the best of the three faecal immunochemical tests are used. An accompanying spreadsheet is provided for self-checking and calculation with different counts from other studies, if such data are available.
**Table 22:** Comparison of sensitive GT with faecal immunochemical tests for different threshold levels of detection for the faecal immunochemical test

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Faecal immunochemical tests (FIT) at different threshold levels (in ng/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitive GT</td>
</tr>
<tr>
<td>Total number of individuals studied</td>
<td>330</td>
</tr>
<tr>
<td>Total Positive based on Colonoscopy</td>
<td>32</td>
</tr>
<tr>
<td>Number of True Positives</td>
<td>17</td>
</tr>
<tr>
<td>Sensitivity of the Test</td>
<td>0.531</td>
</tr>
<tr>
<td>Specificity of the Test</td>
<td>0.594</td>
</tr>
<tr>
<td>Total Negative based on Colonoscopy</td>
<td>298</td>
</tr>
<tr>
<td>True Negative</td>
<td>177</td>
</tr>
<tr>
<td>False Positive</td>
<td>121</td>
</tr>
<tr>
<td>Total Positive by the Test</td>
<td>138</td>
</tr>
<tr>
<td>False Negative</td>
<td>15</td>
</tr>
<tr>
<td>Positivity of the Test</td>
<td>0.418</td>
</tr>
<tr>
<td>Relative Sensitivity</td>
<td>NA</td>
</tr>
<tr>
<td>Increased (or change) in Sensitivity in percent</td>
<td>NA</td>
</tr>
<tr>
<td>Relative Positivity</td>
<td>NA</td>
</tr>
<tr>
<td>Change in Positivity in percent</td>
<td>NA</td>
</tr>
<tr>
<td>PPV for Sensitive GT</td>
<td>0.123</td>
</tr>
<tr>
<td>PPV for FIT</td>
<td>NA</td>
</tr>
<tr>
<td>PPV for FIT/PPV for GT</td>
<td>NA</td>
</tr>
<tr>
<td>If the number of colonoscopy with Hemoccult Sensa is:</td>
<td>4000</td>
</tr>
<tr>
<td>Then, change in the number of colonoscopies if Faecal Occult Blood Tests are used at the specified threshold values of detection</td>
<td>NA</td>
</tr>
</tbody>
</table>

**Discussion**

As can be seen from Table 22, a faecal immunochemical test at threshold levels between 100-125 reduced the need of colonoscopies by about 3000 colonoscopies (or nearly 70%) compared to the ones required by using Hemoccult Sensa in a high risk population. However, there are several limitations of this approach worth indicating:

1. The data came from a study that included very low number of subjects.
   However, this is the only study that was identified that included a head to head...
comparison between individuals who received a faecal immunochemical test versus those who received sensitive guaiac test for screening of advanced adenoma.

2. The data also were relevant to a group of individuals who were at high risk of developing colorectal cancer as opposed to an average risk cohort of individuals. Therefore, figures quoted in this figure are likely to be different if an average risk population was selected.

3. The large difference in the performances of the two tests (even at various levels of threshold of detection) might have arisen due to the low specificity value observed for the sensitive GT.