The following full text is a publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/109531

Please be advised that this information was generated on 2018-01-26 and may be subject to change.
Similar fecal immunochemical test results in screening and referral colorectal cancer

Sietze T van Turenhout, Leo GM van Rossum, Frank A Oort, Robert JF Laheij, Anne F van Rijn, Jochim S Terhaar sive Droste, Paul Fockens, René WM van der Hulst, Anneke A Bouman, Jan BMJ Jansen, Gerrit A Meijer, Evelien Dekker, Chris JJ Mulder

AIM: To improve the interpretation of fecal immunochemical test (FIT) results in colorectal cancer (CRC) cases from screening and referral cohorts.

METHODS: In this comparative observational study, two prospective cohorts of CRC cases were compared. The first cohort was obtained from 10,322 average risk subjects invited for CRC screening with FIT, of which, only subjects with a positive FIT were referred for colonoscopy. The second cohort was obtained from 3,637 subjects scheduled for elective colonoscopy with a positive FIT result. The same FIT and positivity threshold (OC sensor; ≥ 50 ng/mL) was used in both cohorts. Colonoscopy was performed in all referral subjects and in FIT positive screening subjects. All CRC cases were selected from both cohorts. Outcome measurements were mean FIT results and FIT scores per tissue tumor stage (T stage).

RESULTS: One hundred and eighteen patients with CRC were included in the present study: 28 cases obtained from the screening cohort (64% male; mean age 65 years, SD 6.5) and 90 cases obtained from the referral cohort (58% male; mean age 69 years, SD 9.8). The mean FIT results found were higher in the referral cohort (829 ± 302 ng/mL vs 613 ± 368 ng/mL, P = 0.02). Tissue tumor stage (T stage) distribution was dif-
Conclusions: Differences in T stage distribution largely explain differences in FIT results between screening and referral cohorts. Therefore, FIT results should be reported according to T stage.

Key words: Screening population; Referral cohort; Fecal immunochemical test; Tumor stage distribution; Colorectal cancer

INTRODUCTION

Colorectal cancer (CRC) is a disease well suited for population-based screening. In industrialized countries, CRC is one of the three malignancies with the highest incidence and mortality[1,2]. CRC and adenomas can cause lower gastrointestinal bleeding, and early detection by guaiac-based fecal occult blood tests (g-FOBTs) can decrease mortality[3,4]. Recently, fecal immunochemical tests (FITs) have been found to be superior over g-FOBTs[5,6].

Studies on FITs have used different designs, e.g., in terms of populations studied. Screening and referral cohorts each have their pros and cons in this respect[6,7]. Studies in screening cohorts, consisting of individuals with average-risk for CRC, best reflect the true target population. However, in most such studies, only subjects who test positive on FIT are referred for colonoscopy, which means that sensitivity and specificity cannot be determined directly[8]. In addition, the number of cases detected, particularly cancers, is usually low. Other designs include high risk or referral cohorts. Here, an important advantage is that colonoscopy is performed in all patients allowing for calculation of direct sensitivity and specificity[9]. In addition, these studies often yield more cases, allowing for more detailed subgroup analyses[10,11].

It has been suggested that conclusions from referral studies cannot be extrapolated to the screening setting[12,13]. So far, no comparative data have been published to verify or falsify this hypothesis, and arguments both in favor of, as well as against this hypothesis exist. Due to the higher pretest likelihood and presence of symptomatic individuals included in referral cohorts, the risk of work-up bias exists, limiting extrapolation to population-based screening. On the other hand, CRC stage distribution has been shown to influence sensitivity of blood-based CRC markers[14], and is likely to have a large influence on FIT results[15]. Therefore, the aim of the present study was to compare FIT results between subjects with CRC found in either a screening or a referral cohort, and determine if differences can be explained by tumor characteristics.

MATERIALS AND METHODS

Study design

The present study aimed to compare FIT results in CRC cases derived from two methodologically different cohorts. Here, CRC cases with a FIT result ≥ 50 ng/mL from both a screening and a referral cohort were compared for mean FIT result, age, sex and tumor stage. Both these studies were initiated in the same time period in the same country, and both used the same FIT.

Setting

Screening population: The screening cohort in this study was prospectively selected from June 2006 to February 2007 by a randomized selection from a Dutch population. Details from this study are described elsewhere[16,17]. In short, eligible individuals 50-75 years of age were invited by mail to perform either a g-FOBT or a FIT (n = 20,623). Participants with a positive test were offered colonoscopy in two academic centers. Exclusion criteria were institutionalization and bowel symptoms[17]. In the current study, only participants enrolled in the FIT arm (n = 10,322) were included.

Referral population: The referral cohort in this study was derived from a previous and ongoing study on FIT performance, and a more detailed description can be found elsewhere[18]. In short, from June 2006 to October 2009, all ambulatory patients (aged ≥ 18 years) scheduled for elective colonoscopy in five Dutch hospitals (including an academic center and large teaching hospital), were selected regardless of the indication for colonoscopy. Eligible subjects were invited to participate in a prospective study on FIT performance. All participants performed a FIT before bowel preparation. Until June 2008, subjects were invited to perform both a g-FOBT and an FIT[18]. In...
Outcomes and measurements

FIT: In both cohorts, an identical semi-quantitative FIT was used: OC-sensor (Eiken Chemical Co., Tokyo, Japan). No restrictions for diet or medication in the week prior to FIT were given. Participants were educated by illustrated and written instructions to sample their feces, ensuring that no contamination with water or urine occurred. In the screening population, participants were asked not to perform the test if bowel cleansing was incomplete.

The FIT used in this study consists of a sampling tube, filled with stabilizing buffer. Participants were instructed to scrape the probe at different parts of the stool. The amount of feces that can be inserted into the sample bottle is regulated to approximately 10 mg. In the referral population, subjects performed the FIT within 72 h before colonoscopy, and returned the test and informed consent form on the day of colonoscopy. All samples were placed at -5°C on arrival and analyzed according to the manufacturer’s instructions within 1 wk, or were frozen at -20°C. The analyses were performed by two experienced technicians, blinded to the clinical data. In the screening population, participants were instructed to return the test by mail as soon as possible. If the test could not be returned immediately, storage in a refrigerator was advised. Again, two specially trained technicians who were blinded to the clinical data processed all samples.

All FITs were analyzed with the OC sensor MICRO desktop analyzer (Eiken Chemical Co.). The agglutination reaction is dependent on the hemoglobin concentration in the sample. A prozone effect may occur if the concentration is too high and the excess amount of antigen limits agglutination. Measured values can then be higher or lower than the actual concentration in the sample. The risk of the prozone effect gradually increases well above 1000 ng/mL. Therefore, every measurement above 1000 ng/mL was classified as 1000 ng/mL. The quantitative nature of the test was maintained, because 1000 ng/mL is at least 10 times higher than the most usual cut-off values between 50 ng/mL and 100 ng/mL.

Colonoscopy and detected malignancies: Colonoscopies were performed under conscious sedation with midazolam and fentanyl at the discretion of the endoscopist. In both cohorts, all colonoscopies were performed or supervised by experienced gastroenterologists. Colonoscopy was considered complete if the cecum was intubated with visualization of the ileocecal valve or the appendiceal orifice, or by intubation up to CRC. Incomplete colonoscopies were excluded. In addition, subjects were excluded in case of insufficient bowel cleansing, as judged by the individual endoscopist. In the screening cohort, an incomplete colonoscopy was followed by a second colonoscopy with propofol anesthesia. If necessary, a computed tomographic colonoscopy was performed followed by a second colonoscopy. If an incomplete colonoscopy in the referral cohort was followed by a complete second colonoscopy, virtual colonoscopy or x-colon within 6 mo, the results were included in analysis.

In the screening study, histology of tissue samples obtained during colonoscopy was evaluated by one experienced pathologist. In the referral cohort, lesions were evaluated according to routine procedures. In both studies, the outcome variable CRC was classified according to tissue tumor stage (T stage) of the TNM-classification (6th edition) according to the AJCC cancer staging manual.

Statistical analysis

The primary aim of the study was to compare mean FIT scores in CRC cases found in the referral and screening setting, with and without correction for CRC T stage. For analyses, only individuals with CRC and a FIT result ≥ 50 ng/mL were selected, because this cut-off value was used for colonoscopy referral in the screening population.

FIT scores do not follow a Gaussian curve. On average, even after correction for the prozone effect, the curve is considerably skewed to the left. Logarithmic transformation of the FIT scores allowed for using the t-test, as a normal distribution was achieved. Multivariate logistic regression analysis was used to evaluate which variables could explain the differences in FIT scores between CRC patients found in the screening and referral cohorts. In logistic regression analysis, the outcome variable was mean FIT score, and the independent variables were population of origin, T stage, age, and sex. Logistic regression analysis was performed both by forward and backward selection.

Statistical analysis was performed with SAS for Windows, version 8.02 (SAS Institute Inc., Cary, NC). Two-sided P values < 0.05 were considered statistically significant.

Ethical approval and consent

In both studies, informed consent was obtained from all participants. Approval and consent from the screening arm of this study was obtained by the Dutch Health Council (2005/03WBO, The Hague, The Netherlands, www.gezondheidsraad.nl). In all centers participating in the referral arm of this study, local Medical Ethics Review Board approval was obtained prior to the start of the study.

RESULTS

Participants

In the screening population, 10,322 subjects were invited to FIT sampling. Of these, 6,157 completed and returned the test. Five hundred and twenty-six participants were scheduled for colonoscopy because the FIT result was ≥ 50 ng/mL (i.e., positivity rate of 8.5%). In 428 patients, colonoscopy was performed, and considered complete in
van Turenhout ST et al. Screening and referral FIT results

**Figure 1** Study flow of the screening and referral cohort compared in the present study. This includes cases in which an incomplete initial colonoscopy was followed by a second complete colonoscopy, virtual colonoscopy or x-colon. FIT: Fecal immunochemical test; CRC: Colorectal cancer; IBD: Inflammatory bowel disease.

**Table 1** Primary indications for colonoscopy

<table>
<thead>
<tr>
<th>Indication group</th>
<th>Indication for colonoscopy</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symptomatic/suspect</td>
<td>Weight loss</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>Clinical suspicion of diverticulitis</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Clinical suspicion of IBD</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>Abdominal pain</td>
<td>310</td>
</tr>
<tr>
<td></td>
<td>Anemia</td>
<td>174</td>
</tr>
<tr>
<td></td>
<td>Hematochezia</td>
<td>418</td>
</tr>
<tr>
<td></td>
<td>Altered bowel habits</td>
<td>416</td>
</tr>
<tr>
<td>Clinical or radiological suspicion of CRC</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>Colonscopy for polypectomy</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>Diarrhea</td>
<td>115</td>
<td></td>
</tr>
<tr>
<td>Constipation</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1725</td>
<td></td>
</tr>
<tr>
<td>Screening/surveillance</td>
<td>Average risk</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>Familial history of CRC</td>
<td>387</td>
</tr>
<tr>
<td></td>
<td>Lynch syndrome</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Polyp surveillance</td>
<td>491</td>
</tr>
<tr>
<td></td>
<td>Post CRC surveillance</td>
<td>157</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>1146</td>
</tr>
<tr>
<td>Other</td>
<td>Not specified/others</td>
<td>143</td>
</tr>
<tr>
<td>Grand total</td>
<td></td>
<td>3014</td>
</tr>
</tbody>
</table>

Primary indications for colonoscopy among 3014 patients included in the referral arm of this study in a comparison of fecal immunochemical test results in screening and referral colorectal cancer (CRC) cases (exclusion criteria used in the referral arm of the study were age < 18 years, hospitalization, colostomy, inflammatory bowel disease (IBD) or total colectomy).

402 (colonoscopy completion rate 94%; Figure 1). In the 26 cases in which the cecum was not visualized, a second complete colonoscopy was performed. In total, 28 cases of CRC were detected.

In the referral population, 3637 subjects were invited for participation. Six hundred and 23 subjects were excluded because of FIT sampling violating the study protocol, incomplete or contaminated colonoscopy, or missing histology of lesions found. Therefore, 3014 individuals who had a complete colonoscopy and histology remained for analysis (completion rate colonoscopy 91%, Figure 1). The indication for colonoscopy was evaluation of symptoms in 57%, screening and surveillance in 38%, and unspecified in 5% (Table 1). The positivity rate (≥ 50 ng/mL) was 15.7%. In total, 105 subjects with CRC were found, of whom 96 (91.4%) had a positive FIT. The sensitivity of FIT for detection of CRC at cut-off values of 50 mg/mL, 75 mg/mL and 100 mg/mL was 91.4%, 90.5% and 89.5%, respectively. The respective specificity for these cut-off values was 83.7%, 85.7% and 87.0%.

In total, 124 patients with CRC and a positive FIT result were found: 28 derived from the screening population and 96 from the referral population. From the referral population, six cases were excluded because the actual T stage could not be determined due to neoadjuvant radiotherapy or palliative treatment. The mean age of the remaining 90 CRC cases from the referral cohort was significantly higher compared with the 28 cases from the screening cohort; 69 (SD 9.8) vs 65 (SD 6.5) years, respectively (P = 0.04). As expected, the proportion of males was higher in both populations and not statistically significantly different between the referred and screened population (58% and 64%, P = 0.54, Table 1). Other abnormalities that might cause (minor) mucosal bleeding in the colon in addition to CRC, potentially influencing the FIT results were seen in 64% of the referral and 79% (not significant) of the screening population (Table 2).

**T stage distribution and FIT results**

The 28 CRC cases from the screening population had a
mean FIT score of 613 ng/mL (SD 368 ng/mL), which was significantly lower ($P = 0.02$) than the mean FIT score of the 90 CRC cases from the referral population 829 ng/mL (SD 302 ng/mL, Table 3).

The CRC stage distribution was different between the populations: early stages were more frequently found in the screening population (Figure 2). Of the 28 cases from the screening population, 13 cases were classified as T1 (46%), eight as T2 (29%), six as T3 (21%) and one as T4 carcinoma (4%). In the referral population, 12 patients had stage T1 tumor (13%), 22 stage T2 (24%), 52 stage T3 (58%) and four stage T4 (4%).

After stratifying mean FIT scores by T stage, none of the tumor stages had a significant difference in FIT results (Table 2): T1, $P = 0.22$; T2, $P = 0.79$; T3, $P = 0.13$. There was only one T4 case in the screening population, therefore, T4 cases were combined with T3 cases. Again, for the combined T3 and T4 stage category, no significant difference in FIT score was seen between both populations ($P = 0.19$).

**Logistic regression analysis**

By univariate analysis, FIT results from the referral cohort were significantly higher compared with the screening cohort ($P < 0.01$). However, after adding T stage to the model, the difference in FIT results lost statistical significance ($P = 0.10$). The $P$ values per T stage were 0.23 (T1), 0.79 (T2), and 0.11 (T3/T4). By multivariate analysis including the variables sex, age, T, N and M stage, only T stage ($P < 0.001$) and (marginally) age ($P = 0.05$) could significantly explain the differences between the screening and referral cohorts.

**DISCUSSION**

The current study compared FIT results in individuals with CRC from two different prospective study designs, i.e., a population-based screening study and a referral cohort, to study the uncertainty about the link between the results obtained from these kind of studies. Cases of CRC from the screening cohort were found to have significantly lower FIT results compared to those from the referral cohort, but after stratifying for tissue tumor stage, no difference remained. In the screening cohort, 75% (21/28) and in the referral cohort only 38% of cases had a T1 or T2 tumor (34/90). Logistic regression analysis confirmed that not the type of population, but only T stage and to a lesser extent age could explain the differences in FIT results of CRC patients between these screening and referral cohorts.

The results of this study are relevant for the evaluation of CRC screening tests and guidance of future study designs. Several studies in screening populations have been performed comparing, e.g., FIT with g-FOBT\[^{[9,12,21]}\]. However, despite the large number of screened individuals in these studies, the absolute number of individuals with CRC was relatively low, hampering subgroup analysis. Furthermore, colonoscopy was missing in subjects with a negative FIT, impeding calculation of sensitivity and specificity. Indeed, for the investigation of the performance of a screening test like FIT, the ultimate prospective study design would contain full colonoscopic evaluation of all participants. However, in a screening population, this is considered unethical or unfeasible because the capacity and/or funds are lacking. In referral populations, FIT negatives do indeed all undergo colonoscopy, and in addition, in much less time and at a substantially lower cost, a much larger number of CRC patients can be included. This enables, e.g., more elaborate subgroup analysis of early stage CRC. The present study shows that tumor stage is the major contributor to the outcomes of FIT between cohorts. Possible differences in characteristics between the cohorts did not have much influence on FIT outcomes. It is indicated here that, if correction or stratification for CRC tumor stage distribution is applied, both screening as well as referral cohorts can be used to answer several important screening-related research questions. Research questions on accuracy of screening tests with sufficient power, could therefore initially be explored using referral populations. In line with Tao and colleagues, who found that sensitivity of blood-based CRC markers is dependent on tumor stage\[^{[14]}\], results for test characteristics should be presented per tumor stage. By merging results from different sources, the strength of the evidence available will be enlarged.

Some considerations need to be discussed for proper interpretation of the present results. For evaluation of sensitivity and specificity, data from referral studies could be extrapolated, or the incidence of interval cancers could be used as false negatives. The latter requires intensive follow-up over many years before sensitivity can be estimated; time in which tumors may further evolve. Second, the number of screening cases is limited, although comparable with other screening studies\[^{[9,12,21]}\]. This limits the power to determine any existing differences. However, from Figure 2 and Table 2 it is clear that it is unlikely that adding more screening cases could change the results substantially. Supported by the results from other

<table>
<thead>
<tr>
<th>Screening population</th>
<th>Referral population</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n = 28)</td>
<td>(n = 90)</td>
</tr>
<tr>
<td>Male %</td>
<td>64 (21)</td>
</tr>
<tr>
<td>Age (yr, mean ± SD)</td>
<td>65 ± 6.5</td>
</tr>
<tr>
<td>Location of CRC (% left sided)</td>
<td>62 (21)</td>
</tr>
<tr>
<td>CRC only (%)</td>
<td>32 (36)</td>
</tr>
<tr>
<td>CRC including (%)</td>
<td>14 (16)</td>
</tr>
<tr>
<td>Advanced adenomas</td>
<td>16 (57)</td>
</tr>
<tr>
<td>Other adenomas</td>
<td>5 (18)</td>
</tr>
<tr>
<td>Other polyps</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Diverticula</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Hemorrhoids</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

\(^{1}\) \(\chi^2\) test: \(P = 0.54\); \(^{2}\) \(t\) test: \(P = 0.04\). The percentages of other abnormalities in addition to colorectal cancer (CRC) in the referral population totaled 65% due to rounding. The total percentage of CRC cases with additional pathology was 64%.

van Turenout ST et al. Screening and referral FIT results...
The present study aimed to compare test performance in two study designs, each with pros and cons, knowing the essential differences between the two cohorts. It can be concluded that T stage reflects the majority of the differences in overall FIT results between the two studied cohorts.

In conclusion, apparent differences in FIT results between screened and referred CRC patients can be attributed to tissue tumor stage. Based on these findings, we conclude that results from both cohorts could strengthen the evidence available. Using referral populations for studying FIT, and potentially also new CRC screening tests, can be useful to stimulate progress in CRC research. Here, sensitivity and specificity should be studied as these measures are independent of the prevalence of the disease\cite{24}, and test characteristics should be stratified by tumor stage. This will be of particular benefit in research questions that require large numbers of cases or colonoscopy confirmation in all individuals, and do not seek predictive values as outcome.

**ACKNOWLEDGMENTS**

The authors gratefully acknowledge the participants, technical analysts, supporting data managers and the endoscopy units for their contributions to both initial studies; van Oijen MGH, Verbeek ALM, van Krieken HH and the Comprehensive Cancer Center Amsterdam are especially acknowledged for their contribution to the initial screening study; the authors especially acknowledge Loffeld RJLF, Wesdorp ICE, van Heukelum HA, and van Hengel E for their contribution to the original studies from the referral arm of the present study; the OC sensor MICRO desktop analyzer used in the referral study was provided by Eiken Chemical co., Tokyo, Japan.

**COMMENTS**

**Background**

Colorectal cancer (CRC) is a disease well suited for population-based screening. Subjects who test positive on fecal immunochemical tests (FITs) should be referred for colonoscopy. FIT performance has been studied in two different types of populations. In studies in screening cohorts, only subjects who test positive on FIT are referred for colonoscopy. Sensitivity and specificity cannot be calculated directly. In addition, the number of cases detected is usually low. Designs with referral cohorts do not study average but high-risk individuals.

In this study, the existing difference between both cohorts could therefore be even less, because time to analysis in the screening cohort was on average somewhat longer. Finally, in the screening cohort, no information about preoperative radiotherapy was available and therefore tumor stage could have been underestimated in a few cases. In summary, the correction for other potential bleeding sources, time lag to analysis, and radiotherapy would even decrease the differences between the referral and screening cohorts and therefore support our conclusions.

The present study aimed to compare test performance in two study designs, each with pros and cons, knowing the essential differences between the two cohorts. It can be concluded that T stage reflects the majority of the differences in overall FIT results between the two studied cohorts.
Research frontiers

FITs detect occult human blood, what might be derived from adenomas or CRC. FITs are more sensitive than guaiac fecal occult blood tests, by which screening has been shown to decrease CRC-related mortality. However, exact FIT characteristics are the subject of debate.

Innovations and breakthroughs

The present study is the first to compare results obtained from both screening and referral populations to gain insight into the comparability of results derived from both study designs. It was shown that referral populations have a different tumor stage (T stage) distribution compared to screening populations (i.e., a higher percentage of high T stage cancers). This was accompanied by higher mean FIT results. After correction for T stage, mean FIT results were similar in both populations.

Applications

Apparent differences in FIT results between screened and referred CRC patients can be attributed to tissue tumor stage. Results from both cohorts could strengthen the evidence available. Using referral populations for studying FIT, and potentially also new CRC screening tests, can be useful to stimulate progress in CRC research, when test characteristics are stratified by tumor stage.

Peer review

This is an important study comparing the appropriate interpretation of FIT in screening population and in patients referred for colonoscopy regardless of the indication. According to the authors’ conclusion differences in T-stage distribution largely explained differences in FIT results between screening and referral indication. According to the authors’ conclusion differences in T-stage distribution both populations.

Applications

Higher percentage of high T stage cancers). This was accompanied by higher tumor stage (T stage) distribution compared to screening populations (i.e., a higher percentage of high T stage cancers). This was accompanied by higher mean FIT results. After correction for T stage, mean FIT results were similar in both populations.

REFERENCES


S-Editor: Gou SX  L-Editor: Kerr C  E-Editor: Zhang DN