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

Correspondence to: Sietze T van Turenhout, MD, Department of Gastroenterology and Hepatology, VU University Medical Center, PO Box 7057, 1081 HV Amsterdam, The Netherlands. s.vanturenhout@vumc.nl
Telephone: +31-20-4440613 Fax: +31-20-4440554
Received: January 15, 2012 Revised: June 21, 2012 Accepted: June 28, 2012 Published online: October 14, 2012

Abstract

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 3637 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-
different between both populations [screening population: 13 (46%) T1, eight (29%) T2, six (21%) T3, one (4%) T4 carcinoma; referral population: 12 (13%) T1, 22 (24%) T2, 52 (58%) T3, four (4%) T4 carcinoma], and higher T stage was significantly associated with higher FIT results \( P < 0.001 \). Per tumor stage, no significant difference in mean FIT results was observed (screening vs referral: T1 498 ± 382 ng/mL vs 725 ± 374 ng/mL, \( P = 0.22 \); T2 787 ± 303 ng/mL vs 794 ± 341 ng/mL, \( P = 0.79 \); T3 563 ± 368 ng/mL vs 870 ± 258 ng/mL, \( P = 0.13 \); T4 not available). After correction for T stage in logistic regression analysis, no significant differences in mean FIT results were observed between both types of cohorts (\( P = 0.10 \)).

CONCLUSION: 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.

© 2012 Baishideng. All rights reserved.

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

Peer reviewers: Peter Laszlo Lakatos, MD, PhD, 1st Department of Medicine, Semmelweis University, Koranyi S 2A, H1083 Budapest, Hungary; Dr. Paul Sharp, Department of Nutrition, King’s College London, Franklin Wilkins Building, 150 Stamford Street, London SE1 9NH, United Kingdom


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 \[10\].

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 \( \geq 50 \) 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 \[6,16\]. In short, eligible individuals 50-75 years of age were invited by mail to perform either a g-FOBT or a FIT (\( \alpha = 20.623 \)). Participants with a positive test were offered colonoscopy in two academic centers. Exclusion criteria were institutionalization and bowel symptoms \[10\]. In the current study, only participants enrolled in the FIT arm (\( \alpha = 10.322 \)) 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 \[17\]. In short, from June 2006 to October 2009, all ambulatory patients (aged \( \geq 18 \) 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 \[7\]. In
an ongoing study after that period, subjects were invited to perform an FIT only, but twice (on two consecutive days). From the latter cohort, the FIT performed 1 d before colonoscopy was selected for the current study for comparability with the screening population. Exclusion criteria were hospitalization, colostomy, inflammatory bowel disease or total colectomy.

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 blood was visible.

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. 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 [34].

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 \( \geq 50 \) 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 [59], 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) [34]. 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 [3].

RESULTS
Participants
In the screening population, 10,322 subjects were invited to FIT sampling. Of these, 6157 completed and returned the test. Five hundred and twenty-six participants were scheduled for colonoscopy because the FIT result was \( \geq 50 \) ng/mL (i.e., positivity rate of 8.5%). In 428 patients, colonoscopy was performed, and considered complete in
null
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. 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, 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. 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
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, 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, Wetsorp 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. However, as colonoscopy is performed in all subjects direct sensitivity and specificity cannot be calculated. 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, 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.
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 cohorts. Therefore the absolute value of the FIT results should be reported according to T-stage.

REFERENCES
25. van Turenhout ST et al. Screening and referral FIT results