CHAPTER 7

FIT false positives in colorectal cancer screening experience psychological distress up to 6 weeks after the colonoscopy

M.J. Denters, M. Deutekom, M.L. Essink-Bot, P.M. Bossuyt, P. Fockens, E. Dekker

Submitted
ABSTRACT

Purpose: Screening programs for colorectal cancer (CRC) aim at reducing mortality. We assessed psychological effects of being invited to an immunochemical fecal test (FIT)-based screening program.

Methods: Asymptomatic persons aged 50-74 years were invited to a Dutch screening pilot. The psychological consequences questionnaire (PCQ) was used to measure the psychological effects of screening. Screen positives had two additional measurements: before undergoing the colonoscopy and four weeks after receiving the colonoscopy findings.

Results: 3,828 invitees (46% male, mean age 60 years) completed the first PCQ. FIT positives had the highest mean total PCQ score (8.32, SD 8.84; score range 0-36) followed by those who declined participation (3.72, SD 6.30). Participants still waiting for their FIT result had a mean total PCQ score of 2.74 (SD 5.11); those with a negative FIT result had the lowest score: 2.06 (SD 4.43) (p<0.001). 195 FIT positives completed the pre-colonoscopy and 253 the post-colonoscopy questionnaires. Mean total, physical and social PCQ scores, decreased significantly from first questionnaire to pre-colonoscopy but scores on the emotional subscale remained at a similar level. In false positives, mean total, physical and emotional PCQ scores decreased significantly. In true positives, a significant decrease in mean PCQ score was observed for the emotional subscale.

Conclusion: Psychological consequences for invitees to a Dutch FIT based colorectal cancer screening pilot differ depending on timing and FIT result. FIT positives are most distressed. FIT positives who have a normal colonoscopy still experience psychological distress 6 weeks after the colonoscopy.
INTRODUCTION
Screening programs for colorectal cancer (CRC) aim at detecting CRC in an early stage and thereby decreasing CRC-related morbidity and mortality. As is the case with any screening program, only a small proportion of persons will directly benefit from participation, whereas a large proportion of invitees have to undergo screening. Invitees are confronted with the possibility of having cancer, which can cause distress by itself.

Most of what is known on adverse psychological effects following screening stems from studies among women invited for breast cancer screening. These studies focused mainly on adverse effects associated with receiving false positive mammography results, since these women are exposed to distress surrounding a positive screen but do not benefit in terms of an early detection of significant lesions. A recent meta-analysis of 17 studies demonstrated that receiving a false-positive mammogram was associated with greater anxiety and distress about breast cancer. Negative psychological effects can persist for up to three years after the screening process.

To our knowledge, only two studies have addressed negative psychological effects or distress in screening for CRC using stool tests. These Scandinavian studies demonstrated no substantial adverse effects on psychological well-being 12 months following CRC screening.

We designed a study to assess the screen-specific psychological effects of being invited for and/or participating in a pilot-CRC screening program using the fecal immunochemical test (FIT). We aimed to explore differences in psychological effects cross-sectionally after receipt of the invitation in all invitees, and longitudinally in participants with a positive test result.

METHODS
Data were collected in the second round of a Dutch biennial FIT-based CRC-screening pilot in the Amsterdam region. The program design has been reported in detail elsewhere. A summary is given below.

Design of the second round of the screening pilot
A random sample of average risk persons aged 50 to 74, living in the screening pilot catchment area in the Amsterdam region of the Netherlands, was selected from the population database based on date of birth and postal code. They were sent an invitation package for a second screening round two years after the first round.

Invitation
The package included an invitation letter, an information leaflet, a FAQ card and the stool test (OC-Sensor by Eiken, Tokyo, Japan; single test at one occasion with a hemoglobin value of 50ngHb/ml as threshold for test positivity) with a detailed test instruction. In the invitation letter, invitees were instructed not to take part in screening but to contact their general practitioner in case of bloody stool and/or changed bowel habits during the last three months. The package further included information on the screening procedure, the
meaning of a positive and a negative test result, the possibilities of false positive and false negative test results, and a figure illustrating the chances of having a positive test result and of being diagnosed with advanced adenomas or CRC at the follow-up procedure. The information leaflet was designed specifically for this study based on literature review and feedback from the first round. It was put together in close relation with a linguistic expert specialized in patient information provision.

**Screening procedure**

Invitees could participate in the screening trial by performing the FIT at home and returning the test in a postage free envelope. They received the results within two weeks after performing the test. Participants with a negative test result were informed through a letter by postal mail. The message specified that the stool sample did not contain blood and that follow-up investigations were unnecessary at this time. The letter explained that a negative test result does not guarantee complete absence of significant lesions and emphasized that persons should contact their general practitioner in case of any alarming symptoms suggestive of cancer, such as blood in stool, or changed bowel habits. In case of a positive test result, screening participants were informed through a letter that the test result was ‘unfavourable’ meaning that blood was detected in their stool sample. The blood could be indicative of cancer but could also come from other, less serious sources, such as haemorrhoids, benign polyps, or fissures.

Participants with a positive test result were invited for a visit to the screening center. In the absence of contraindications they were scheduled for colonoscopy within 2 weeks. On the day of the colonoscopy, participants were informed about the preliminary results before going home. Two weeks after the colonoscopy all participants were contacted by phone or by face-to-face consultation to discuss the histopathological results and follow-up.

**Psychological consequences of screening questionnaire (PCQ)**

Psychological consequences of screening were elicited with a questionnaire at three different time-points. We sent a first questionnaire two weeks after the invitation (the FIT-questionnaire). To distinguish between participants that had already been informed on their test result and participants that were still waiting, we included two additional items in the first questionnaire, asking whether or not participants had already performed the test and whether or not they had been informed about the results.

We sent a second questionnaire to participants with a positive test result just before the colonoscopy procedure (the pre-colonoscopy questionnaire) and a third questionnaire to all positive screenees that underwent a colonoscopy four weeks after they had been informed about the colonoscopy results, which is about six weeks after the colonoscopy had taken place (the post-colonoscopy questionnaire).

To evaluate psychological consequences of screening we used the Dutch version of the Psychological Consequences of Screening Questionnaire (PCQ). This is a validated measure on screen-specific anxiety, originally designed for breast cancer screening. The PCQ consists of 22 items. Since the focus of our study was on adverse psychological effects we only used the 12 items on the negative consequences of screening.
In the PCQ invitees are asked to indicate how often they experienced each of a list of 12 symptoms ‘over the past week as a consequence of thoughts and feelings about colorectal cancer’. Answers could be given on a 4-point Likert scale anchored at 0 (not at all) and at 3 (quite a lot of the time). The 12 items referred to the effect of screening on an individual’s functioning on emotional (e.g. feeling down, feeling nervous, worried about future), social (e.g. holding back, having trouble going to work/meeting others) and physical life (e.g. sleep disturbances, changed appetite) domains. Each domain represents a subscale for which a total score can be calculated. The PCQ total score can range between 0 and 36, with the emotional subscale ranging between 0 and 15, the social subscale between 0 and 9 and the physical subscale between 0 and 12. Higher scores indicate more adverse effects.

Data analysis
We included data from all invitees that had returned the first questionnaire in the analysis. We analysed the data by time point and by test result. Primary outcome measure was the PCQ total score. Secondary outcome measures were the PCQ emotional, physical and social subscores.

Data for the first time point were analysed conditional on the responses to the question whether or not they had performed the FIT, and, if so, whether or not they had already been informed about the test result, and, if so, on the type of test result. For those who had indicated they had been informed about the test results we retrieved the result from the screening database. Hemoglobin levels of over 50ngHb/ml were classified as a positive test result and haemoglobin levels under 50ngHb/ml as negative. This way we could classify respondents into four groups: invitees who had not performed the test at the time of completing the first questionnaire, invitees who had performed the test but who were still waiting on their results, invitees who had performed the test and had received a negative test result, and persons who had performed the test and had received a positive test result. For each group we computed a mean PCQ total score at the first timepoint. We compared means between groups using the ANOVA test statistic. We hypothesized that scores would be different, with participants who had already been informed about a negative test result having the lowest score and participants who had been informed about a positive test result the highest score.

We then analysed the scores at the second time point, prior to colonoscopy. We computed a mean total PCQ score of all FIT positives who had returned the pre-procedure questionnaire and compared it with the scores of the subgroup that was already aware of their positive FIT result at the first timepoint. We hypothesized that the pre-colonoscopy PCQ total score would be higher than the first PCQ score in this subgroup. To test for a significant change in PCQ total score, we included all FIT positive participants with a both a baseline PCQ total score and a pre-procedural PCQ total score to compute a paired t-test statistic.

For our analysis of the third time point, the post-colonoscopy PCQ scores, we organized participants into two groups, based on their colonoscopy results: persons with a true positive FIT result and persons with a false positive FIT result. Persons with a true positive FIT result were those in whom at least one carcinoma or advanced adenoma had been detected at colonoscopy. Advanced adenomas were adenomas sized 10mm or larger or adenomas with a villous component of more than 20% or high-grade dysplasia. All other participants having undergone the colonoscopy were classified as false positives. We hypothesized that
the post-colonoscopy PCQ score would be higher for true positives than for false positives and tested those using ANOVA. To test for significant changes in PCQ score before and after colonoscopy, we included all FIT positive participants with both a pre-colonoscopy and a post-colonoscopy PCQ score to compute paired t-test statistics.

A significance level of 0.05 was used in all hypothesis tests. Data were analysed using the statistical software SPSS 18.0.

Ethical approval
Ethical approval for the screening program and data collection was provided by the Dutch Health Council (2005/03WBO, The Hague, The Netherlands).

RESULTS
A total of 10,265 persons were invited for participation in the second screening round (49% male, mean age 60±7). Of these, 5,367 (52%) returned the FIT and 424 (8%) of participants had a positive test result. Most test positives (n=373; 88%) underwent a follow-up colonoscopy, of which 163 (44%) had a true positive test result and 210 (56%) a false positive test result.

First questionnaire
Overall, 3,828 invitees completed the first PCQ questionnaire. Table 1 shows the characteristics of questionnaire responders and non-responders. As can be appreciated from this table, males were less likely to return the questionnaire and questionnaire responders were on average older than non-responders. Of 3,828 invitees who returned the first PCQ, 228 (6%) had not performed the test at the time of questionnaire completion, 1,385 (36%) had performed the test but were still waiting on their test result, 2,053 (53%) had already received a negative test result and 162 (4%) had already received a positive test result.

Figure 1a-d show the mean PCQ total scores as well as the mean scores for the emotional, physical and social subscale scores for the four groups specified above on the first questionnaire. The mean PCQ total score at the first time point was highest in participants who had received a positive test result (mean total PCQ score 8.32, SD 8.84, 95% CI 6.95 to 9.69), followed by those who had not performed the test (mean PCQ 3.72, SD 6.30, 95% CI 2.90 to 4.54) and participants who were still waiting on their result (mean PCQ 2.74, SD 5.11, 95% CI 2.47 to 3.01). The mean PCQ total score was lowest in those who had received a negative test result (mean PCQ 2.06, SD 4.43, 95% CI 1.87 to 2.25). All differences were statistically significant (p<0.001). Similar patterns were observed for the three PCQ subscales: overall group differences in mean score at the first time point were significantly different (p<0.001). Post-hoc analysis showed no significant difference between the ‘no test performed’-subgroup and the ‘waiting for test result’-subgroup on the PCQ physical and PCQ social subscales.

Pre-colonoscopy analysis in FIT positives
Of 373 FIT positives who underwent a colonoscopy, 195 (52%) completed the pre-colonoscopy questionnaire. Their mean pre-colonoscopy PCQ total score was 6.86
Table 1. Age and gender of questionnaire responders and non-responders.

<table>
<thead>
<tr>
<th></th>
<th>Responders (n=3,828)</th>
<th>Non-responders (n=6,437)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males, n (%)</td>
<td>1,374 (46)</td>
<td>3,635 (50)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Mean age ±SD, years</td>
<td>60.3 ±6.9</td>
<td>59.5 ±6.9</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Figure 1a-d. a. Changes in mean PCQ total score (score range 0-36). b. Changes in mean PCQ emotional subscore (score range 0-15). c. Changes in mean PCQ physical subscore (score range 0-12). d. Changes in mean PCQ social subscore (score range 0-9). Figures show the mean score in function of time. Error bars represent the standard error of the mean.
(SD 6.78, 95% CI 5.91 to 7.82), significantly lower than the mean score in those who had learned of a positive FIT result at the time of first questionnaire completion (p= 0.020). The mean PCQ score on the physical and social subscales also showed a significant decrease (p=0.017 and p=0.010 respectively) but scores on the emotional subscale remained at a similar level (p=0.31).

**Post-colonoscopy analysis comparing true positives and false positives**

Of the 373 FIT positives who underwent a colonoscopy, a total of 253 (68%) persons completed the post-colonoscopy questionnaire: 133 (53%) with a true positive FIT result and 112 (47%) with a false positive result. We observed no significant differences between false positives and true positives in mean post-colonoscopy PCQ scores for the PCQ total nor for any of the PCQ subscale scores (Figure 1). Analysis of the changes in PCQ score over time showed that in false positives the mean PCQ total and mean PCQ physical and emotional subscale scores were significantly lower after colonoscopy – reflecting fewer adverse psychological consequences - whereas the mean PCQ social score remained at a similar level (p=0.32). In true positives, a significant decrease in mean PCQ score was observed for the emotional subscale only (p=0.026).

**DISCUSSION**

We observed that being invited for FIT-based CRC-screening was associated with different levels of psychological distress depending on the FIT result and the time point in the screening process. Adverse psychological effects of participation were most pronounced shortly after having been informed about a positive test result. Just before undergoing the colonoscopy, the overall level of psychological distress was significantly reduced. We argue that this might be due to the consultation at the outpatient clinic that participants had attended in the mean time. During this consultation the meaning and possible consequences of positive test result were elaborately discussed. Nevertheless, the scores on the emotional subscale remained at an increased level just before undergoing the colonoscopy.

Six weeks after the colonoscopy had taken place, overall psychological distress was lower compared to pre-procedure in both true en false positives.

Strengths of this study are the large number of persons that completed the questionnaire and the fact that it was conducted among a true screening population. The design of the questionnaire allowed us to compare four different groups of invitees at baseline: non-participants, participants still waiting on their result and persons already aware of their result. A limitation is the low response rate among non-participants. A second limitation is that we cannot exclude selection bias. It is possible that invitees who experienced high levels of psychological distress were more likely to return the questionnaire. We also do not have PCQ scores from a reference population unexposed to screening for comparison. We can only compare differences in psychological consequences within invitee subgroups exposed to the screening invitation. Finally, since screen positives were only followed up once after colonoscopy at six weeks, we do not how scores develop over a longer period of time.
We are aware of two studies that looked at psychological effects of being invited for cancer screening using stool tests. Both studies used other measures making direct comparison difficult. A recent Danish study evaluated the psychological distress following fecal occult blood test (FOBT) screening among average risk persons invited for participation in a feasibility study. They found that at baseline, FOBT positive participants were more likely to be worried of having CRC and stated a higher degree of concern about participating in the actual screening program. They also had significantly higher somatization, depression and anxiety scores compared to FOBT negatives. After 3 months only anxiety scores remained significantly higher in FOBT positives and at 12 months no significant differences in any of the domains remained.

Lindholm and colleagues assessed distress created by a stool test based CRC-screening program in participants and non-participants by a combination of a questionnaire and a telephone survey. They found that 16% of participants and 15% of non-participants reported severe distress after having received the invitation letter. In around 40% of both participants and non-participants, this distress lasted for more than a week. Surprisingly, the effects of this distress on daily life were more pronounced for non-participants. The observation that non-participants experienced more effects on their daily life than participants is similar to our finding that non-participants experience higher levels of psychological distress than persons still waiting on their test result and persons who already received a negative test result. These findings suggest that merely receiving an invitation to participate in screening can cause severe distress in approximately 2 out of 10 invitees. Another explanation could be that non-participants are not participating in screening because of fear.

In Dutch cervical cancer screening, Korfage and colleagues have used the PCQ to assess the psychological distress in women with low-grade abnormalities in their Pap-smear and compared these to a reference group of cervical screening participants awaiting smear taking. The mean total PCQ score was 5 in the group with abnormalities. This score was slightly lower than the score observed in FIT positives in our population (8.32). A Swiss study that also assessed psychological consequences of a false positive mammography result showed results more similar to ours. They assessed the level of psychological consequences in screen positives 8 weeks after the notification that no abnormalities were found at follow-up. These women had a mean emotional score of around 3, a PCQ social score of 1 and a PCQ physical score of 2. When we compare these numbers to the PCQ scores we observed in false positives 6 weeks after the notification that no abnormalities were found at colonoscopy, we observed fairly similar scores.

In conclusion, this study and the data described above show that being invited for participation in a screening program can cause psychological distress in invitees depending on the test result and the time point in the screening process. Although some level of distress seems inevitable after receiving a positive screening result, we also showed that persons in whom ultimately no significant lesions were found (i.e. the false positives) still experienced distress six weeks post-colonoscopy. We should be aware of these adverse effects and try to minimize them as much as possible, especially in those persons who do not directly benefit from screening in terms of early detection of (pre)cancerous lesions. Future studies should explore the evolution of these adverse effects to answer the question whether in the case of CRC screening these scores remain at a higher level or return to baseline in, for example, 12 months time.
REFERENCE LIST


