Distressed or relieved? Psychological side effects of breast cancer screening in the Netherlands


Abstract

Study objectives—To assess the psychological impact of mammographic screening on women with non-malignant outcomes after attending the Netherlands’ National Breast Cancer Screening Programme.

Design—During one year all women with false positive test results (95) in a screening area were invited for the study. Each false positive was matched with two women with normal mammograms with respect to age and municipality. A random reference group of 400 was drawn from the female population in an area not yet included in the screening programme. Experiences with screening and psychological status of subjects were assessed 8–10 weeks after screening (T1) and again after six months (T2), by interviews as well as questionnaires. References completed two questionnaires with a six months’ interval.

Participants—74 (78%) women with false positive outcomes and 113 (59%) women with negative outcomes participated at T1, of these 65 (88%) and 105 (93%) at T2, respectively; 238 references returned questionnaires at T1 (59%), of these 143 (60%) at T2.

Main results—At 8–10 weeks after the screening, the women who received false positive test results scored higher on most of the variables indicating psychological disfunctioning than women with normal mammograms, but did not notably differ on the same variables from the non-screened reference group. Women with normal mammograms had the lowest scores on all the variables in the study at both assessments. The same situation was observed six months later. Although 61% of the women who received false positive mammograms reported that they had experienced the “false alarm” as a stressful event, this experience had apparently no adverse effects on their psychological functioning, as assessed 8–10 weeks after screening.

Conclusions—Overall, breast screening is not likely to generate adverse psychological effects in “healthy” women, even if the outcome is false positive. Differences in psychological functioning between false positives and negatives are more likely ascribable to feelings of relief in the negative group than to raised anxiety and distress in the false positive group.

Not only will the number of women who may benefit from regular mammography increase by the diffusion of population based screening, but also the number of women who may be affected by psychological side effects. Side effects can arise from different stages of the screening process, from the invitation for mammography confronting women with the possibility of having breast cancer to the diagnosis of a breast cancer that was not suspected. The comparatively few studies on the subject suggest that the psychological impact of breast screening followed by a negative outcome is negligible or non-existent in “psychologically healthy” women, but that a false positive outcome can cause emotional disturbances up till six months or more after being informed that the abnormal mammogram did not imply breast cancer. Women with false positive outcomes and subsequent clinical examination were reported to have suffered more from distress, anxiety, and breast cancer worries than women with negative outcomes. In most cases the severity of these emotions subsided after some time (that is, the difference between false positives and negatives had disappeared) and, so far, persistent psychiatric morbidity resulting from breast screening has not been assessed.

However, by the existing evidence the conclusion that screening for breast cancer has no adverse psychological effects may be somewhat premature. Raised anxiety levels, more worries about breast cancer than usual, and feelings of insecurity after a “false alarm” should be considered negative side effects, even if they do not result in psychiatric morbidity and occur “only” for a limited period. Such feelings can affect the daily functioning and might have repercussions on the adherence to future mammographies. Now that regular screening mammography is promoted as normal preventive health behaviour of women in the target population, more understanding is needed of the impact of this procedure on “psychologically healthy” women.

The aim of this study was to evaluate the psychological consequences of accepting an invitation for mammography within the scope of a national breast screening programme, both for women who received false positive results and for women who received negative results.

Methods

PROCEDURES

The Netherlands’ National Breast Screening Programme offers free biennial mammography to 50–69 year old women. The overall...
attendance is 78% at the initial screening and 73% at subsequent screening rounds. Participants are informed of their test results by mail within two weeks after the mammogram. The current rate of abnormal mammograms in the study region is 1.6%, half of which are diagnosed false positive and half of which positive after clinical examination.  

The comparatively small rate of false positive test results is an inconvenient condition for creating an optimal study design. A prospective design would require a very large baseline sample to provide for a satisfactory number of false positives in the study. For example, to acquire data of 75 false positives in a follow up study, 12 500 baselines are needed presuming a response of 75%. For this study, data were collected retrospectively and the results of the screened subjects, both false positives and negatives, were compared with those of a random reference group of non-screened women.

The subjects were included in the study after attending their initial mammography in a screening area in the north of the Netherlands, the inclusion period was one year. The attendance at first screen in this area is 82%. All women with false positive outcomes were invited (n=95). These women were matched on a 1–2 basis (n=190) with negatives of the same age and of the same municipality. Both groups were asked to participate in two semi-structured interviews. A false positive and her matched negatives were interviewed within a period of four weeks. The first series of interviews (T1) were conducted as soon as possible after the outcome of the clinical examinations of the false positives, which was, because of administrative procedures and rules on privacy of medical data, 8–10 weeks after the screening. The second series (T2) were conducted six months after the first. Experienced female interviewers gave the interviews in the respondent’s own house. At the end of each interview, a short questionnaire was left behind, containing self-report measures to assess psychological functioning. Respondents were expected to complete this questionnaire and to return it (prepaid) by mail. A random community sample was drawn from the female population of 50–69 years in an adjacent region that was not yet included in the breast screening programme (n=400). These references were not interviewed but received two identical postal questionnaires with an interval of six months, containing all the self-report scales and most of the structured questions that were used in the study. The first batch of questionnaires was sent about five months after the start of the T1 interviews of the subjects.

CONTENTS OF INTERVIEWS AND QUESTIONNAIRES
The interviews started with open ended questions allowing respondents to elaborate on their experiences with the screening process proper, the clinical examination (false positives), and their feelings in this respect. The aim was to gain a deeper insight in how women, of whom the majority are unsuspecting, react to a "false alarm" of breast cancer; information that could not probably be obtained by structured questions or health status instruments only. The psychological status of subjects and references was assessed by standardised instruments and forced choice questions. The instruments had to be brief, comprehensible, and easy to complete by respondents of all educational levels. At the same time they had to be effective in assessing comparatively mild emotional disorders in non-psychiatric people. The following instruments were used:

Mood disturbances
The 12 item General Health Questionnaire (GHQ-12) was used as an overall index of distress, both as a continuous variable and as an instrument to identify "cases" and "non-cases". The scoring scale ranges from 0–12. A threshold score of 2 was chosen to include cases with comparatively mild to moderately severe disorders.  

The Hospital Anxiety and Depression scale (HAD) was used for differentiating between anxiety and depression. An advantage of this scale is that it specifically excludes items referring to somatic symptoms that might be attributable to physical illness. The scoring scale per subscale of seven items ranges from 0–21. The HAD distinguishes between "non-cases" (threshold score 0–7), "doubtful cases" (8–10), and "definite cases" (11 or more).  

Somatisation
As psychosomatic complaints are often indicative of beginning or mild mood disturbances, the 12 item somatic complaint sub-scale (SOM) of the Symptom Checklist-90 was included as a dependent variable; scoring scale: 12–60. Added are two single item Likert style questions about sleep disturbances and loss of appetite, two aspects not measured by the SOM sub-scale (item scores 1–5).

Cancer related variables
General fear of cancer was assessed by an eight item Dutch validated "Fear of Cancer Scale".  

The scoring scale ranges from 8–32. Additional indicators of anxiety, referring specifically to breast cancer, were worries about breast cancer during the past two weeks (before the assessment), the frequency of breast self-examination
Table 1: Background characteristics of respondents at T1 by screening outcome and reference groups

<table>
<thead>
<tr>
<th></th>
<th>False positive (n=74)</th>
<th>Negative (n=113)</th>
<th>Reference (n=238)</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>57.9</td>
<td>57.7</td>
<td>59.2</td>
<td>0.04</td>
</tr>
<tr>
<td>Never married</td>
<td>71.6</td>
<td>84.7</td>
<td>79.4</td>
<td></td>
</tr>
<tr>
<td>Widowed</td>
<td>9.5</td>
<td>5.4</td>
<td>3.4</td>
<td></td>
</tr>
<tr>
<td>Divorced</td>
<td>16.2</td>
<td>9.0</td>
<td>10.7</td>
<td></td>
</tr>
<tr>
<td>Living alone</td>
<td>2.7</td>
<td>0.9</td>
<td>6.0</td>
<td></td>
</tr>
<tr>
<td>Formal education of respondent*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elementary training</td>
<td>70.8</td>
<td>60.4</td>
<td>64.5</td>
<td>0.29</td>
</tr>
<tr>
<td>Secondary training</td>
<td>23.6</td>
<td>25.2</td>
<td>26.8</td>
<td></td>
</tr>
<tr>
<td>College, university and advanced professional training</td>
<td>5.6</td>
<td>14.4</td>
<td>8.6</td>
<td></td>
</tr>
<tr>
<td>Formal education of partner*</td>
<td></td>
<td></td>
<td></td>
<td>0.08</td>
</tr>
<tr>
<td>Elementary training</td>
<td>60.7</td>
<td>51.6</td>
<td>46.3</td>
<td></td>
</tr>
<tr>
<td>Secondary training</td>
<td>28.6</td>
<td>25.3</td>
<td>37.3</td>
<td></td>
</tr>
<tr>
<td>College, university and advanced professional training</td>
<td>10.7</td>
<td>23.2</td>
<td>16.4</td>
<td></td>
</tr>
</tbody>
</table>


(BSE) during the past four weeks, and the degree of reassurance after the diagnosis that no evidence of cancer was found. Breast cancer worries were assessed by a single five point Likert style item, previously used in other studies on the psychological effects of false positive mammograms; subjects rated their current levels of breast cancer worry from "not at all" (1) to "most of the time" (5). Frequency of BSE in the past four weeks ranged from 0 to 4 times or more.

Demographic characteristics
These included age, marital status, living status (single or not), and educational levels of respondents and their partners.

SUBJECTS AND RESPONSE
Seventy four women with false positive outcomes participated in the first series of interviews (78% of 95 invited) and 113 women with negative outcomes (59% of 190 invited). Of these, 65 false positives (88%) and 105 negatives (93%) also participated in the second series (T2). Response on the first postal questionnaire in the reference group was 59%, 60% of these responders returned the second questionnaire.

Table 1 summarises the demographic characteristics of the respondents at T1. Women in the reference group were a little older than the screened subjects and, compared with the false positives, more women in the negative and reference groups were married. In other respects the three groups were similar. Non-respondents at the second assessment (T2), both subjects and references, did not differ from respondents with respect to demographic characteristics.

STATISTICAL ANALYSIS
Statistical analyses were performed, using Pearson’s χ² statistic on categorical data and univariate analysis of variance (oneway) on continuous data, group means were compared by the Student-Newman-Keuls procedure. Multivariate analyses of variance (MANOVA) were conducted to test the effect of outcome status on the dependent variables simultaneously, this to preclude "capitalising on chance" and to control for differences in demographic characteristics. Intra-group differences between the two measurement points were tested by paired Student’s t tests. p Values are based on two tailed tests and results are considered significant if p < 0.05.

Results
PSYCHOLOGICAL AND PSYCHOSOMATIC STATUS
Table 2 shows the mean scores and standard deviations of univariate analyses at T1.

Mood
Eight to 10 weeks after the screening, no differences were found in distress or depression between women with false positive outcomes, women with negative outcomes, and references. Anxiety in the false positive group was higher than in the negative group, but not significantly different from the references; negatives had lower anxiety scores than references.

Somatisation
The somatic complaints sub-scale of the SCL-90 did not discriminate between false positives, negatives or references, but women in the false positive group complained more of appetite and sleep than those in the negative and reference groups.

Cancer related variables
Women who were false positive were more afraid of cancer in general than women who were negative, but the fear levels in both groups

Table 2: Psychological and psychosomatic status by screening outcome and reference groups at T1

<table>
<thead>
<tr>
<th></th>
<th>False positive (FP) (n=74)</th>
<th>Negative (NEG) (n=113)</th>
<th>Reference (REF) (n=238)</th>
<th>Comparison of means*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean score</td>
<td>SD</td>
<td>Number†</td>
<td>Mean score</td>
</tr>
<tr>
<td>Mood</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distress (GHQ-12)</td>
<td>1.71</td>
<td>3.11</td>
<td>65</td>
<td>0.82</td>
</tr>
<tr>
<td>Depression (HAD)</td>
<td>2.80</td>
<td>3.93</td>
<td>65</td>
<td>2.54</td>
</tr>
<tr>
<td>Anxiety (HAD)</td>
<td>4.29</td>
<td>3.68</td>
<td>66</td>
<td>2.93</td>
</tr>
<tr>
<td>Somatisation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Somatic complaints (SCL-90)</td>
<td>17.95</td>
<td>7.02</td>
<td>70</td>
<td>16.63</td>
</tr>
<tr>
<td>Loss of appetite</td>
<td>1.35</td>
<td>0.78</td>
<td>74</td>
<td>1.02</td>
</tr>
<tr>
<td>Sleep disturbances</td>
<td>2.38</td>
<td>1.34</td>
<td>74</td>
<td>1.61</td>
</tr>
<tr>
<td>Cancer related variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fear of cancer</td>
<td>13.34</td>
<td>8.66</td>
<td>65</td>
<td>11.15</td>
</tr>
<tr>
<td>Breast cancer worries</td>
<td>1.38</td>
<td>0.83</td>
<td>73</td>
<td>1.07</td>
</tr>
<tr>
<td>Frequency of BSE</td>
<td>1.27</td>
<td>1.35</td>
<td>74</td>
<td>0.94</td>
</tr>
</tbody>
</table>

* p < 0.05.
† Numbers refer to women who supplied complete data.
of screened women were significantly lower than in the population sample. False positives also more often worried about breast cancer than the negatives in the two weeks before assessment, but as with anxiety, not more than women in the reference group. No differences were found in frequency of BSE between the three groups.

It is notable that the lowest mean scores on all the dependent variables in the study were found in the negative group.

A similar pattern can be observed at the second assessment (T2), six months after the first (table 3). Negative women still had the lowest mean scores on the variables in the study and this time the false positives did not differ from the references on any of the variables. At T2 the negatives differed from false positives and references on distress, but not on anxiety. Somatic complaints as defined by the SCL-90 sub-scale were now significantly more frequent in the false positive group compared with the negative group and the differences in appetite or sleep disturbances found at T1, had for the most part disappeared.

Non-response at T2, intra-group changes between the two assessments
To look for selective response at the second assessment, the T1 scores of the non-responders at T2 were compared with the T1 scores of the subjects who responded. In the false positive group no systematic differences were found. However, in the negative group, the eight non-responders at T2 had higher mean T1 scores on both distress (p < 0.10) and anxiety (p < 0.05), whereas the non-responders at T2 in the reference group had lower distress scores at T1. No differences were found between responders and non-responders at T2 with respect to the remaining variables in the study.

For those subjects who supplied complete data at both measurements the intra-group changes between the two assessments were analysed.

The only significant changes in mean scores on the dependent variables observed between T2 and T1, were a decrease in complaints about appetite and sleep in the false positive group. Otherwise, the changes within each of the three groups between the two assessments were only small and insignificant.

**Psychiatric morbidity**
No significant differences were found in prevalence of GHQ cases ≥ 2 at T1, but the GHQ case rate at T2 was lower for negatives than for false positives or references (table 4). Of the 18 cases at T1 in the false positive group, 12 (67%) had remained cases at T2, three (17%) had become non-cases, and three (17%) had incomplete data in the self report questionnaires or were non-respondents at T2. In the negative group, four (22%) of 18 cases at T1 were still cases at T2, 10 (56%) had become non-cases, and four (22%) had incomplete data or were non-respondents. Because of the small numbers of subjects who could be defined as HAD cases, borderline and definite cases were combined. Prevalence of HAD anxiety ≥ 8 was lower in negatives than in false positives or references both at T1 and T2. Six of 11 anxiety cases in the false positive group at T1 were also cases at T2, four were non-cases, and one was a non-respondent. In the negative group two of seven cases at T1 were cases at T2, one had become a non-case, and four were non-respondents or supplied incomplete data. Prevalence of HAD depression cases ≥ 8 did
not differ at both assessments between the three groups.

RESULTS OF MULTIVARIATE STATISTICAL ANALYSIS

Separate MANOVAs were conducted on data of each assessment. Outcome status (false positive, negative, reference) was treated as an independent variable and those variables that were significantly associated with outcome status in the univariate analyses at either T1 or T2, were treated simultaneously as dependent variables. The results revealed a significant effect of the screening outcome on the psychological variables in the model (Hotellings $T^2$, $p < 0.0001$) both at the first and at the second assessment. This effect was the result of lower mean scores on the dependent variables of the screened negative group (see table 2 and 3) and it remained significant after adding age and marital status to the models as confounding variables.

RESPONDENTS' OWN ACCOUNT OF THEIR EXPERIENCES

In the open ended, “in depth” part of the interview the false positives were stimulated to elaborate on their emotional reactions to the (mail delivered) message that they were referred for clinical examinations and on their experiences afterwards. At the time of the first interview, 8–10 weeks after the screening, 61% of the false positives told that they had responded strongly to the notification of their referral. Adjectives used were “stunned”, “shocked”, “utterly upset”, “panic-stricken” or “very angry”. Sixteen per cent of all false positives were almost certain they had breast cancer, 40 per cent said that they felt very frightened although they did not actually mention “cancer” or “the worst”. These reactions occurred although they were informed, both in the invitation leaflet and the referral letter, that abnormalities on the mammogram do not automatically imply breast cancer. Fifty seven per cent of the false positives had not reckoned at all with the possibility of an unfavourable outcome of the screening. In this respect they did not differ from the women whose outcome was negative. The period between the notice of referral and the conclusion of the clinical examination with a favourable outcome was described as a very anxious and trying period by 32 per cent of the false positives. At the time of the second interview the acuteness of the emotions about the “false alarm” had subsided in most of the women concerned. Of the 65 false positives who responded at T2, 8% reported that they still thought very often about what had happened to them, 27% sometimes, and 65% only occasionally. Seventy three per cent felt totally reassured by the outcome of the clinical examination, 20% fairly reassured, and 6% was only a bit reassured or not at all. By comparison, 93% of the women who were negative felt totally reassured by the outcome and 7% fairly reassured.

Despite their experiences, most of the false positives (89%) who did respond at T2 thought positive about breast screening in general and intended to participate again in future rounds. Two false positives were sure that they would not participate because of unpleasant experiences, and five women were not sure because the event had made them very anxious. Eight of nine false positives who refused a second interview had in the first interview expressed strong negative feelings concerning the “false alarm”. Overall 28% of the respondents in the false positive group saw themselves as “sensible people”, not to be easily upset. They described their first reactions to the abnormal mammogram as “keep calm, because nothing is certain yet”. They reported that they had rationally and open mindedly undergone all examinations in the hospital. The clinical diagnosis of “no breast cancer” was according to their expectations and they experienced no negative effects. For them the case was closed with the outcome of the clinical examinations.

Discussion

With respect to the differences between false positives and negatives, the results of this study are very much consistent with those of other studies that investigated the psychological impact of “false alarm” after breast screening. Eight to 10 weeks after the screening, false positives were more anxious, more afraid of cancer, and more worried about having breast cancer than women who had normal mammograms. False positives also complained more about appetite and sleep. At first sight these results confirm the assumption that a false positive test result and the delayed assurance of not having cancer is a stressful event, which can have negative psychological consequences for the women concerned, at least for some time. However, the introduction of a random sample of non-invited women from the general population puts things in a different light. It seemed that the psychological status of the false positives did not differ much from that of the non-screened population sample, but that the women who were negative had lower scores on all the indicators used. One other study considered the prevalence of breast cancer anxiety in screened women and in the general population. That study showed a significantly lower level in negatives and a slightly, but not significantly, higher level in false positives compared with the population sample. Another interesting aspect of the present study is the fact that in the qualitative part of the interviews a comparatively large proportion of the false positives reported that their referral for clinical examination had come as a shock, partly because they had hardly reckoned with the possibility of anything but a “normal” outcome, and that they had gone through a very stressful and anxious time in their life. Apparently these experiences did not result in (moderately) psychological ill effects. It is conceivable that the peak of the emotional disturbances occurred in the first few weeks after the event and had levelled off at the time the subjects could be contacted, which was not earlier than 8–10 weeks after the screening. If so, it can be assumed that women in general can manage...
the tension and anxiety that are provoked by breast screening.

A reference group of non-screened women is a legitimate alternative to a baseline measurement in studies where using a baseline is not feasible. The advantage of this study is the inclusion of a comparatively large number of false positives. The few prospective studies on the subject included only a few or no respondents with false positive screening outcomes.\(^1\) A disadvantage is the attrition at the second assessment, specifically in the reference group.

In psychosocial research where subjects are recruited by introductory letters or postal questionnaires response problems are not uncommon and one of those things investigators have to cope with. The response rates at T1 are satisfactory for all three groups and comparable with those of others.\(^2\)\(^3\)

However, for two of the dependent variables there are indications of a selective response in the negative and the reference groups at the second assessment.

The question is to what extent this affects the results of the study. The response at the first assessment does not give reasons to doubt the general picture emerging from the results beyond the usual care with respect to differences from sampled data.

As far as a response bias can be presumed for the second assessment, this regards two variables in the negative group and one variable in the reference group. Had the non-responders remained in the study and had their levels on the concerning variables remained stable over time, both the mean scores on distress and anxiety in the negative group at T2 might have been slightly higher and the mean score on distress in the reference group lower. Consequently, the differences of the false positives with the negatives might have been somewhat smaller and with the references larger. As there are no indications of a response bias in the false positive group, possible divergencies from the observed values at T2 are more likely to affect the conclusions for the negative and the reference groups.

We have not found any evidence of a selective response in connection with the other variables in the study, for example, those referring to “cancer anxiety” or breast cancer worries. So, although the results of the second assessment have to be treated with caution, we assume that our study extends the evidence that breast screening is not likely to have adverse psychological effects on “healthy women”, even if the outcome is false positive. On the contrary, it seems like breast screening followed by a “normal” outcome may have a positive effect on the well being of the women who participate, at least for the first two months after the screening. With respect to fear of cancer in general and breast cancer worries it seems that such an effect may last even longer.

Another point that deserves attention is the fact that the results of the study are based on group means. It is, therefore, not impossible that a small percentage of women may suffer from prolonged anxiety or cancer consciousness, or both, because of the screening process. Furthermore, the comparatively large standard deviations of the dependent variables in the false positive group do not exclude the possibility of a small subgroup of women with high scores. It is also possible that women who experienced serious problems resulting from the screening are underrepresented in the study. For the present, the relative size of such a group appears rather small and does not give reasons to introduce standard intervention programmes or counseling services for women who received a false alarm. First, more knowledge is needed of the specific nature of the problems and the way to identify the women at risk. Acquiring such knowledge might not be an easy task because of the small numbers of the women involved.

The apparent effect of relief observed in the negative group, specifically at the first assessment, is an interesting finding that also deserves further research.

The authors are grateful to The Northern Netherlands Comprehensive Cancer Society and the Northern Netherlands Breast Cancer Programme (director H H Meehroth) for their kind cooperation. They put much effort into providing data and introduced the researchers to the women eligible for the study. We would also like to thank our research assistant Mrs G Holander and Dr E Van Sondersen, who gave valuable advice on data management and statistical analysis, and of course we are very much indebted to our competent interviewers and to all the women who participated in this study. Mrs A Bos-Nijenhuis corrected the English text of the manuscript.

Funding: the study was funded by The Dutch Cancer Society (grant no 93-138).

Conflict of interests: none.