Classification of findings in mammography screening—a method to minimise recall anxiety?

Martti Pamilo, Jouko Lönnqvist, Arja Halttunen, Martti Soiva, Ilkka Suramo

Abstract

Study objective—The aim was to find out if it is possible, by classifying screening mammograms according to the likelihood of malignancy, to divide the recalled women to a group in which there is high suspicion of malignancy, most having breast cancers, and a group with more obscure findings.

Design—Screening mammograms of recalled women were classified according to the likelihood of malignancy. 0 = technically insufficient, 1 = normal, 2 = benign tumour, 3 = malignancy cannot be excluded, 4 = strongly suspicious for malignancy, 5 = malignant.

Setting—This study was a population based survey of mammography screening in Helsinki and surroundings in Finland.

Patients—21 417 women (aged 50–59 years) were invited to be screened, 18 012 (84.10%) participated. Of these 579 (3.21%) of those screened) were recalled for further studies; 124 of these were referred for surgical biopsy and 82 had breast cancer.

Measurements and main results—All cases classified as 5, 60% of the cases classified as 4, 65% of the cases classified as 3, 0% of the cases classified as 2 or 1, and 1.2% of the cases classified as 0 proved to have breast cancers. However classification 5 represented 5.9% of all recalled women and 41.5% of all screening detected breast cancers; classification 4, 6.0% of all recalled women and 25.6% of all screening detected breast cancers; classification 3, 68.9% of all recalled women and 31.7% of all screening detected breast cancers; classification 2, 11.7% and classification 1, 2.9% of all recalled women. No breast cancers were detected with these classifications. Classification 0 represented 4.5% of all recalled women and 1.2% of all screening detected breast cancers. Classifications 5 and 4 represented only 11.9% of all recalled women but 67.1% of all screening detected breast cancers.

Conclusions—By classifying screening mammograms according to the likelihood of malignancy, recalled women can be divided into two groups: (1) a quite small subgroup in which everyone or almost everyone will be shown to have breast cancer; and (2) a much larger subgroup in which only a few will be proven to have breast cancer. The invitation procedure for the further studies should be improved on this basis of minimising anxiety among recalled women.

Mammography screening has become common and is already a part of normal health services in Great Britain and Finland. Several large local screening studies are in progress and/or under evaluation, most of them showing decreased mortality in breast cancer.

Screening mammography has, however, been criticised for being non-specific and non-sensitive. Although the sensitivity and specificity of well organised mammography screening has been shown, with significantly more breast cancers than benign findings among those operated, some women have to be recalled for detailed studies who do not actually have breast cancer.

Primary screening mammograms are not sufficient for excluding breast cancer in 2 to 5% of cases. These have to be recalled for more detailed examinations. Most of these, however, prove to be normal or benign findings, and breast cancer is eventually found only in about 0.5% of these screened. Evidence indicates that the former group, without malignancy, do not make up a risk group for breast cancer. Thus, for every 100 000 women screened, about 4500 are recalled who will have no signs of breast cancer in further studies.

Regardless of these encouraging figures, screening has psychological costs for those who receive bad news or false positive results. The distress is hard to alleviate even when further investigations are negative. This phenomenon has been noted in various kinds of screening studies (for hypertension, cervical cancer, abnormal folate, and congenital hypothyroidism).

Our experience with Finnish screenings shows that a recall for further studies almost always arouses strong initial anxiety. Even the primary screening arouses anxiety that manifests itself in the attitudes of the refusers. Little is, however, known about the scale or duration of the screening aroused anxiety but what is known is reassuring.

The aim of this study was to classify those recalled into different groups according to the likelihood of malignancy: a group with a high susceptibility for malignancy, most having breast cancers, and group with more obscure findings, most of whom in further studies will have normal or benign findings. This information from the primary screening could be used when recalling the women for further studies by informing them more exactly about the reasons for recall. With this method we predict that it is possible to reduce anxiety markedly among the large majority of those recalled.

Methods

The material comprised the screening mammograms performed at the mammography...
screening centre of the Cancer Society of Finland in Helsinki in the years 1986–88. Over 21 000 women (aged 50–59 years) were invited to be screened, all from Helsinki and 22 surrounding municipalities; 18 012 (84 10−6) participated. Of these, 579 (3.21%) of those screened were recalled for further studies (more detailed mammography, ultrasonography, cytology, and clinical examination). All the women recalled participated. Of the recalled women, 124 (0.69%), of those screened) were referred for surgical biopsy, and 82 (0.46%, of all those screened) had breast cancer.17 The invitation to the free of charge screening mammography was in written form, designed to be as relevant and motivating as possible without arousing anxiety. The wording of this invitation was: “The purpose of the examination is to find a possible breast cancer in an early stage before it can be felt in order to improve the results of treatment”, with the added reassurance: “If you are recalled for further studies, it must be realised that the most of the findings on mammograms will prove to be benign”.

To keep the period of uncertainty and possible anxiety as short as possible, the staff analysed the screening mammograms as soon as possible and sent out the results within one week, usually within 2–3 days after taking the films. They performed the required further studies without delay.

The entire process was as follows: two mammographic projections (mediolateral oblique and craniocaudal) were interpreted separately by two radiologists, and doubtful or suspicious mammograms were interpreted together by the two radiologists after which decisions were made as to whom to recall. After this double reading, the findings were classified: 0 = technically insufficient, 1 = normal, 2 = benign tumour, 3 = malignancy cannot be excluded, 4 = strongly suspicious for malignancy, 5 = malignant. Those classified 3–5 were recalled for further studies, as were those classified as 0. When the classification was 1 or 2, only those furnishing anamnestic information about a palpable lump or bloody/serous nipple discharge were recalled, together with those with benign changes if it was obvious from the size and localization of the finding that it must be palpable in spite of the lack of anamnestic information. One of the two radiologists performed the further studies, after which the final radiological finding was determined as well as the necessity for surgical biopsy.

### Results

The screening results are presented in table I. Table II shows changes in classification of the primary screening findings (0–5) after further studies were performed. In addition, the proportion of surgical biopsies and breast cancers is shown for each group. The primary screening classification involved 5.9% (34/579) of those recalled. All these were referred for surgical biopsy after the further studies, and breast cancer was histologically verified in 32. Two women refused surgery. One of these had both radiologically and cytologically verified malignancy and the other radiologically verified malignancy: typical casting type intraductal microlcifications situated in one lobe and intraductal tumour growth in the same lobe in galactography (performed because of bloody nipple discharge). If these two are also considered as verified breast cancers, all the cases classified as 5 proved to be malignant. Classification 5 represented 41.5% of all screened detecting breast cancers.

Cases with classification 4 represented 6.0%, (35/579) of all recalled women. Of these, 8.6%, (31/35) were referred for surgical biopsy, and 60.0% (21/35) were verified histologically to have breast cancers. Classification 4 represented 25.6%, of all screened detecting breast cancers. Two thirds (67.1%, (22/34)) of screening detected breast cancers was classified as 4 or 5, but these represented only 11.9% (69/579) of all recalled women. Cases in which the primary screening classification was 3 (malignancy cannot be excluded) constituted 68.9% (399/579) of all recalled women. After the further studies, 13.8%, of these were referred for surgical biopsy. Breast cancer was verified in only 6.5%, (26/399) with this primary classification. However these represented 31.7% of all screening detected breast cancers. With the further studies, 65.7%, (262/399) were proved to be superimpositions and/or normal, and 20.3% (81/399) benign lesions. If the primary screening classification was 2 or 1 (11.7%, and 2–9%, of those recalled) no breast cancers were detected.

![Table I Screening results 1986-1988.](image)

<table>
<thead>
<tr>
<th>Invited</th>
<th>21 417</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participated</td>
<td>18 012</td>
</tr>
<tr>
<td>of those invited</td>
<td>84.10−6</td>
</tr>
</tbody>
</table>

| Recalled | 579 |
| Surgical biopsy | 124 |
| Breast cancer | 82 |
| Benign | 42 |

<table>
<thead>
<tr>
<th>Primary screening classification</th>
<th>Radiological classification after further studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Operation</td>
<td>B</td>
</tr>
<tr>
<td>C</td>
<td>n</td>
</tr>
<tr>
<td>0</td>
<td>26</td>
</tr>
<tr>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>68</td>
</tr>
<tr>
<td>3</td>
<td>399</td>
</tr>
<tr>
<td>4</td>
<td>35</td>
</tr>
<tr>
<td>5</td>
<td>34</td>
</tr>
</tbody>
</table>

†Two refused surgery. One had radiologically and cytologically verified breast cancer, the other radiologically verified breast cancer.

B = benign, M = malignant, C = classification (0–5), n = number of cases.
The primary screening classification 0 constituted 4.5% of all those recalled, but only one woman with this classification had a breast cancer (1.2% of all screening detected breast cancers).

Discussion
During each screening round a certain proportion of women are always recalled for further studies. At the beginning of a nationwide screening programme in Finland in 1987 (screening mammography with two projections every year) it was estimated according to Swedish experience that the proportion of recalled women would be 5%, of those screened. That means that when all Finnish women aged 50–59 years are actively included in screening, there will be about 7000 women to be recalled annually. Only about 10% (700) of these women will, however, be referred further for surgical biopsy and only half of these (350) will prove to have breast cancer. When evaluating the usefulness of the 0–5 classification for the primary screening findings, one can see that in classification 5 all 34/34 proved to have breast cancers. When the classification was 4, 21/35 (60.0%) proved to have breast cancers. The other four classifications accounted for scattered cases, with none found in classes 1 and 2 (figure).

The results of our study could be used to reform an invitation policy to minimise anxiety among those women recalled for further studies. At initial screening women should be carefully informed that the two views per breast used in screening mammography are not enough for everybody, so about 2–5% of screened women have to be recalled for further studies. There could be two kinds of blank for recall, one sent to those with strong suspicion of malignancy (classification 4–5) and the other to those with uncertain or equivocal findings that must be supplemented with additional views or other procedures, as well as for those with anamnestic information about a palpable lump or nipple discharge although screening mammograms showed benign findings (classification 0–3). The purpose of this dual recall system would be to minimise anxiety in the latter cases. At least two kinds of arguments can be presented against our suggested invitation policy: it might put those women recalled with the implication of strong suspicion into a more difficult situation psychologically if further studies show no signs of malignancy. This group would, however, be very small compared to those suffering from the current practice in which everybody recalled may respond negatively. One can, of course, send a blank implying strong suspicion only to women in classification 5. Anxiety caused by recall has also a positive motivating effect; this was reflected in the 100% attendance rate for the further studies, whereas in the primary screening 16% of those invited refused. One may fear that the compliance percentage may decrease among women receiving the recall blank implying uncertainty in the screening findings, but we do not consider this probable.

In principle the effects of screening can be health supporting and/or anxiety provoking. The effects which the recall and the further studies inspire can be examined as a process of change, in which an invitation or a recall means a sudden loss of emotional balance in respect to health.

Using the knowledge obtained by further studies, one tries to reach a new balance. Between these events there is, however, a period characterised by uncertainty and anxiety. To furnish correct information on health as quickly and accurately as possible is the main principle of a controlled procedural change.

In minimising the anxiety provoked by the recall and by the further studies we can use mainly cognitive means to support adaptation by giving as detailed and accurate information as possible. Unfortunately, in practice the possibility of supporting individuals directly and also emotionally is very scanty in screenings. However, when evaluating the cost-effectiveness and risk-benefit ratio of screenings one has to consider also the significance of psychological reactions aroused by the invitation and by the actual screening procedure, although no excess psychological morbidity has been shown. Good training of the radiologists and quality control of the screening are, however, the primary methods to improve the specificity of screening.

In conclusion, when interpreting primary mammograms at screening, cases are actually divided into two groups: (1) normal cases, no sign of a breast cancer, and (2) those needing further studies. If mammography findings in the latter group are classified according to susceptibility to malignancy one can then make two subgroups: (1) a small subgroup in which everyone or almost everyone will be shown to have breast cancer, and (2) a much larger subgroup in which only a few will be proven in further studies to have breast cancer.

The invitation procedure should be improved so that with proper information the anxiety could be minimised among those women who are recalled but have no breast cancer.


