False-positive results in mammographic screening for breast cancer in Europe: a literature review and survey of service screening programmes

Solveig Hofvind, Antonio Ponti, Julietta Patnick, Nieves Ascunce, Sisse Njor, Mireille Broeders, Livia Giordano, Alfonso Frigerio and Sven Törnberg The EUNICE Project and Euroscreen Working Groups (Members of the EUNICE Project and Euroscreen Working Groups listed at end of paper)

Objective To estimate the cumulative risk of a false-positive screening result in European mammographic screening programmes, and examine the rates and procedures of further assessment.

Methods A literature review was conducted to identify studies of the cumulative risk of a false-positive result in European screening programmes (390,000 women). We then examined aggregate data, cross-sectional information about further assessment procedures among women with positive results in 20 mammographic screening programmes from 17 countries (1.7 million initial screens, 5.9 million subsequent screens), collected by the European Network for Information on Cancer project (EUNICE).

Results The estimated cumulative risk of a false-positive screening result in women aged 50–69 undergoing 10 biennial screening tests varied from 8% to 21% in the three studies examined (pooled estimate 19.7%). The cumulative risk of an invasive procedure with benign outcome ranged from 1.8% to 6.3% (pooled estimate 2.9%). The risk of undergoing surgical intervention with benign outcome was 0.9% (one study only). From the EUNICE project, the proportions of all screening examinations in the programmes resulting in needle biopsy were 2.2% and 1.1% for initial and subsequent screens, respectively, though the rates differed between countries; the corresponding rates of surgical interventions among women without breast cancer were 0.19% and 0.07%.

Conclusion The specific investigative procedures following a recall should be considered when examining the cumulative risk of a false-positive screening result. Most women with a positive screening test undergo a non-invasive assessment procedure. Only a small proportion of recalled women undergo needle biopsy, and even fewer undergo surgical intervention.

INTRODUCTION

Mammographic screening aims to detect breast cancer at an early stage in women without symptoms. Suspicious findings on screening mammography need further investigation. The decision to recall a woman for further assessment should be a consensus decision made between two radiologists or through arbitration of a third radiologist, and is an essential part of the screening process.

Women who have further assessments and are diagnosed as not having breast cancer are referred to as having a ‘false-positive screening result’, which is considered an adverse effect of mammographic screening. The likelihood of having breast cancer will be higher in women who are recalled for further assessment compared with those who are not, and the likelihood is even higher in women who have a needle biopsy (fine needle aspiration cytology, FNAC, or core needle biopsy, CNB) compared with those who do not. It is therefore not surprising that a recall for further assessment could be stressful.

The psychological impact of being recalled for further assessment in mammographic screening has been reported in several studies. Different psychological measurement scales and quality-of-life assessments have been used and reviews have concluded that false-positive screening results did not cause anxiety and distress at a general level, but rather breast cancer specific distress, anxiety and apprehension. The negative impact on a woman’s wellbeing is seen during the examination period, in the period between screening and further examination, and after further examination, regardless of the final outcome. The adverse effect among women found not to have breast cancer has been
reported to be transient,4,6,8 but occasionally can be seen more than a year after the screening test.3,9

The psychological harm of a false-positive screening result could lead to a lower participation in future screening rounds, but results from studies have been inconclusive.4,7,10–13 A recent meta-analysis from four European countries based on data from more than 340,000 women showed that a false-positive screening result did not have any material influence on the overall participation during subsequent rounds.13

The risk of a false-positive screening result is strongly positively correlated with the recall rate. This rate is influenced by the training and experience of the radiologist, by the image quality and number of views, by other factors related to screening (e.g. prevalence or incidence screen, screening interval, single versus double reading and screening technique) and characteristics of the women (e.g. age, screening history, use of hormone therapy, breast density, previous invasive procedure and familial breast cancer).14–19

To optimize the balance between benefits and harms in mammographic screening, it is important to keep the false-positive rate low without missing breast cancers.20 We reviewed European studies that estimate the cumulative risk of a recall for further assessment among women without a diagnosis of breast cancer, associated with participation in biennial screening mammography starting at ages 50–51 and continuing to ages 68–69. This age group was chosen to be in line with the European guidelines for quality assurance in breast cancer screening and diagnosis.20 We stratified the cumulative risk by whether invasive procedures (needle biopsy and/or surgery) were used. In an additional analysis, we present cross-sectional observed rates of further assessment, with and without needle biopsy, and surgery, and the positive predictive value (PPV) of the screening test for several European countries, based on data collected as a part of the European Network for Information on Cancer project (EUNICE).21

**METHODS**

**Cumulative risk of a false-positive screening result**

Procedures for further assessment following a recall differ between countries and centres, but usually include additional mammographic imaging (additional views and magnification), ultrasound and clinical breast examination as the first step. If the finding on the screening mammograms is not resolved after this step, a needle biopsy (CNB or FNAC) is usually performed to obtain a histological or cytological diagnosis. If the biopsy is positive for cancer, surgical treatment is considered. If no histological or cytological diagnosis is available using needle biopsy, women are usually referred for surgical intervention (surgical biopsy and/or surgical treatment). Preoperative breast MRI is being performed with increasing frequency, particularly after an unsuccessful needle biopsy. These three steps (imaging, clinical assessment and biopsy) represent a typical patient flow associated with procedures for further assessment following a recall, and are crucial for understanding the differences in the described extent and psychological consequences of further assessments with a benign outcome. The number of false-positive screening results includes recalls for further assessment with a final diagnosis that is not breast cancer, regardless of which procedures were performed.

We defined the cumulative risk of having a false-positive screening result as the risk of being recalled for further assessment at least once during 10 biennial screens performed from age 50 to 51 until 68 to 69, among women without a diagnosis of breast cancer. The cumulative risk of an invasive procedure with benign outcome was defined as the risk of having at least one needle biopsy or surgical intervention performed during the same time span, again among women without a diagnosis of breast cancer.

We conducted a literature review based on articles in PubMed with a title, abstract or keywords including a reference to both false-positive results and screening mammography which were published in 1995–2011 (the search was performed in October 2011). The search terms were (false-positive OR abnormal OR benign) AND (breast cancer OR mammog”). We also manually searched the reference sections of relevant papers. We aimed to identify reports of the cumulative risk of a false-positive screening result in women aged 50–69. After searching on title and abstract, and the whole paper if it appeared relevant, four original research papers10,22–24 and one research letter25 were considered relevant for our analysis. The Spanish Cumulative False-Positive Risk Group has published several studies on this topic, based on the same women.18,24,26,27 We chose the first published study, of more than 1.5 million women, including 251,275 aged 50–51 and followed for six screening rounds.24 The 8502 women in the study by Castells et al.22 were also included in Salas et al.24 and the study by Castells et al. was therefore excluded from the pooled analyses and from our review results.

We estimated both the unweighted average of the cumulative risk of a false-positive screening result after 10 screening examinations and the average weighted by study size. The cumulative risk of women undergoing needle biopsy is reported only in two studies.10,24 The cumulative risk of surgical intervention was given only in the study by Hofvind et al.10 As a consequence of the limited numbers of studies covering needle biopsy and surgical intervention, the average rates, both weighted and unweighted, should be considered with care.

**Performance parameters in European mammographic screening programmes**

We augmented the cumulative estimates with cross-sectional estimates from population-based mammography programmes in Europe. Aggregate data regarding performance parameters achieved in European mammographic population-based screening programmes were collected within the EUNICE project in 2008–2009. The EUNICE project is described in detail elsewhere in this Supplement.21 We included data from 20 national or regional screening programmes (from 17 countries), based on 7,658,586 screening examinations performed between 2005 and 2007. Compared with the report on screening coverage and participation,21 seven programmes (two national and five regional) were excluded because of missing
information on key variables concerning further assessment or because data from initial and subsequent examinations could not be separated. One regional programme (MaMMa Network, Budapest) was added.

Information about further assessment procedures following a positive screening test (i.e. needle biopsy and surgical intervention) were calculated as proportions of screened women. The PPV of the screening test was calculated as the number of screen-detected breast cancers (including ductal carcinoma in situ) divided by the number of all positive screening tests (among those with and without cancer). The benign-to-malignant (B/M) biopsy ratio, as well as the rate of surgical intervention includes all procedures performed as a consequence of a positive screening test. Time trends by age groups were analysed using Poisson regression. Spearman’s rank correlation-co-efficient (rho) was used for examining correlations between outcome parameters. Time trends and correlation coefficients were only investigated for subsequent screening tests. In order to ensure that the assumption of linearity was reasonable, the correlation between the rate of further assessment and 1/PPV of the screening test was plotted, the latter of which is interpretable as the number of women recalled per cancer detected.

RESULTS

Cumulative risk of a false-positive screening result

Table 1 describes the studies included, and their respective estimates. Together, they are based on 390,000 screened women.

The study by Hofvind et al.10 was based on women in the first three screening rounds in the pilot study of the Norwegian Breast Cancer Screening Programme, conducted between 1996 and 2001. The cumulative risk of having one or more false-positive screening results was estimated for women aged 50–51, who are expected to be screened 10 times until age 68–69. The authors assumed independence between the outcomes of subsequent screening tests which may overestimate the cumulative risk. To extrapolate to a follow-up time of 20 years, they further assumed that the risk of having a false-positive screening result from each of the fourth to the 10th screening tests was the same as that observed in women in corresponding age groups in the third screening round. It was estimated that women aged 50–51 who participate in 10 biennial screening tests have a 20.8% cumulative risk of a false-positive screening result. The cumulative risks of having FNAC or CNB were estimated to be 3.9% and 1.5%, respectively. The cumulative risk of having a surgical intervention (both surgical biopsy and final surgical treatment) with benign outcome was 0.9%. This was the only study that presented the cumulative risk of surgical intervention with benign outcome.

The largest study (more than 250,000 women) came from Spain.24 The cumulative risk of a false-positive screening result was estimated for women aged 50–51 at the start of screening. The database included women who were screened at least once in eight of the 17 administrative regions of the mammographic screening programme. The estimates were obtained from a regression model (discrete time hazard model), after adjustment for possible confounding factors (screening unit and calendar period). The overall cumulative risk of having a false-positive result was 20.4%, and 1.8% for an invasive procedure (fine or thick needle biopsy, biopsy and/or other invasive tests) with benign outcome.

The study by Njor et al.23 was based on two areas within Denmark. The study used the same estimation model as used by Hofvind et al. but Njor et al. first analysed whether there was independence between the outcomes of subsequent screens. The hypothesis of independence between the outcomes of subsequent screens was accepted, and the estimated cumulative risk of having a false-positive screening result was 15.8% for women residing in the Copenhagen area and 8.1% for women residing in the Fyn area. The risks were 22.6% for Copenhagen and 9.9% for Fyn when crude proportions of participants with a false-positive test from each of the first five invitation rounds were used.23

Puliti et al.25 estimated the cumulative risk of a false-positive screening test at 15.2% in 28,500 women who had participated in seven screening rounds in Florence, Italy (Table 1). The cumulative risk of a needle biopsy with benign outcome was 1.8%.

We derived pooled estimates of the cumulative risks using those studies that estimated the risk over 10 years, that is all except the study by Puliti et al.25 Thus, the pooled estimates were based on 364,991 screenees. The weighted pooled estimate of the cumulative risk of having at least one false-positive screening result over 10 biennial screening sessions in women aged 50–51 without breast cancer was 19.7% based on three studies.10,23,24 The unweighted average was 16.5%. Both are consistent with the empirical result from Italy (15.2% after seven screening rounds). The weighted pooled cumulative risk of having a false-positive screening test that led to an invasive procedure was 2.9% based on two studies (unweighted average 4.1%).10,24 The estimated cumulative risk of a false-positive screening result without an invasive procedure was therefore 16.8%.

Performance parameters collected in the EUNICE project

The EUNICE project is described by Giordano et al.21 in this supplement of the journal. Data from 20 mammographic screening programmes in 17 European countries illustrate that the rate of further assessment varied between 2.2% and 15.6% for initial screening tests, and between 1.2% and 10.5% for subsequent screening tests in women aged 50–69 (Table 2). The overall rates were 9.3% and 4.0%, respectively, for initial and subsequent screening tests. The rate of needle biopsy varied between 0.8% and 3.3% for initial screening tests and between 0.3% and 1.5% for subsequent screening tests respectively (2.2% and 1.1% overall). Surgery was performed in an average of 1.0% of the initial screens (range: 0.4–1.4%) and 0.73% (range: 0.3–1.1%) of the subsequent screens. The variation in further assessment rate is reflected in the PPV which differed substantially in programmes that provided data: 4.9–24.2% in initial screening tests and 6.8–49.5% in subsequent tests. The rate of surgical intervention with benign outcome was only 0.19% in initial and 0.07% in subsequent screening.
Table 1 Summary of studies performed in Europe to estimate cumulative risk of a further assessment without cancer diagnosed during a given screening period

<table>
<thead>
<tr>
<th>Authors and year of publication</th>
<th>Country</th>
<th>Study group (n)</th>
<th>Age group (years)</th>
<th>Study period (years)</th>
<th>Screening interval (year)</th>
<th>Views (#)</th>
<th>Adjusted for characteristics</th>
<th>Observed screening rounds (#)</th>
<th>Screening rounds in estimation (#)</th>
<th>Cumulative risk of further assessments among women without cancer diagnosed</th>
<th>Invasive proc (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hofvind et al. 2004&quot;</td>
<td>Norway</td>
<td>83,416†</td>
<td>50–69</td>
<td>1996–2001</td>
<td>Two</td>
<td>Two</td>
<td>No No</td>
<td>3; values for 4th–10th rounds based on 3rd round</td>
<td>10</td>
<td>20.8</td>
<td>FNAC: 3.9, CNB: 1.5, SI: 0.9, total: 6.3</td>
</tr>
<tr>
<td>Njor et al. 2007&quot;</td>
<td>Denmark</td>
<td>21,261†, 9039†</td>
<td>50–69</td>
<td>1993–2003, 1991–2001</td>
<td>Two</td>
<td>Pre: two, Sub: one; two if dense breast</td>
<td>No No</td>
<td>5; values for 6th–10th rounds based on 5th round</td>
<td>10, 10</td>
<td>8.1, 15.8</td>
<td>NA, NA</td>
</tr>
<tr>
<td>Salas et al. 2011&quot;</td>
<td>Spain</td>
<td>251,275</td>
<td>50–69</td>
<td>1990–2006</td>
<td>Two</td>
<td>One or two</td>
<td>Yes Yes</td>
<td>6; values for 7th and 10th rounds based on 6th round</td>
<td>10</td>
<td>20.4 (20.0–20.8)</td>
<td>1.8 (1.7–1.9)</td>
</tr>
<tr>
<td>Puliti et al. 2011&quot;</td>
<td>Italy</td>
<td>28,500†</td>
<td>50–69</td>
<td>1991–2006</td>
<td>Two</td>
<td>Two</td>
<td>No No</td>
<td>7; empirical data</td>
<td>7</td>
<td>15.2</td>
<td>IP: 1.8</td>
</tr>
<tr>
<td>Weighted pooled estimates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>19.7, 10, 23, 24</td>
<td>All 2.9, 10, 24</td>
</tr>
</tbody>
</table>

All programmes performed independent double reading and assumed independence between subsequent screening tests
FNAC: fine needle aspiration cytology, CNB: core needle biopsy, SI: surgical intervention, IP: Invasive procedure (fine or thick needle biopsy, biopsy and/or other invasive tests)
†Fyn
†Copenhagen
§Ini: initial screening round
§Sub: subsequent screening rounds
C: Crude rates
### Table 2

Number of screening examinations in women aged 50–69 who were screened within 2005–2007 in European mammographic screening programmes, rate (%) of further assessment (FA), preoperative needle biopsy and surgical intervention (all and with benign outcome, including all surgical procedures performed as a consequence of a positive screening test) among those screened, positive predictive value (PPV) of the screening test and benign-to-malignant (B/M) ratio of surgical biopsies, by screening history (initial and subsequent screening test) and area.

| Period       | Screening exams (n) | FA (%) | Needle biopsy (%) | All (%) | Ben* (%) | PPV* (%) | B/M* (ratio) | Screening exams (n) | FA (%) | Needle biopsy (%) | All (%) | Ben* (%) | PPV* (%) | B/M* (ratio) |
|--------------|---------------------|-------|-------------------|--------|----------|----------|----------------|---------------------|-------|-------------------|--------|----------|----------|----------------|            |
| Belgium      | 2005                | 47,104| 6.2               | –      | 0.8      | 0.13     | 10.7          | 0.20                | –     | 0.5               | 0.06   | 14.0     | 0.12     | 0.12          |
| Flanders     |                     |       |                   |        |          |          |                |                     |       |                   |        |          |          |               |
| Czech Republic | 2005–2006          | 256,425| 15.6              | 1.1    | 0.7      | –         | –             | –                   | 234,900| 10.5             | 0.6    | –        | –        | –             |
| Denmark      | 2005                | 2910  | 4.3               | –      | 1.3      | 0.27     | 24.2          | 0.27                | –     | 10.77             | 0.04   | 14.0     | 0.12     | 0.07          |
| Copenhagen   |                     |       |                   |        |          |          |                |                     |       |                   |        |          |          |               |
| Estonia†     | 2005–2006          | 20,555| 2.2               | 0.9    | 0.4      | 0.06     | 15.6          | 0.18                | –     | –                 | –      | –        | –        | –             |
| Finland      | 2005                | 15,631| 7.0               | 2.2    | 0.8      | 0.12     | 7.3           | 0.23                | –     | –                 | –      | –        | –        | –             |
| Hungary      |                     |       |                   |        |          |          |                |                     |       |                   |        |          |          |               |
| Budapest‡‡  | 2005                | 170,427| 7.2              | 1.2    | 0.8      | 0.16     | 7.5           | 0.29                | –     | –                 | –      | –        | –        | –             |
| Italy        | 2005                | 170,427| 7.2              | 1.2    | 0.8      | 0.16     | 7.5           | 0.29                | –     | –                 | –      | –        | –        | –             |
| Luxembourg   | 2004–2005          | 5094  | 14.1              | –      | 0.9      | 0.22     | 4.9           | 0.31                | –     | –                 | –      | –        | –        | –             |
| Norway       | 2005–2006          | 76,058| 5.7               | 2.1    | 0.7      | 0.12     | 9.7           | 0.22                | –     | –                 | –      | –        | –        | –             |
| Poland       | 2007                | 403,596| 10.6              | –      | 1.3      | –         | –             | –                   | –     | –                 | –      | –        | –        | –             |
| Portage centre | 2005              | 13,841| 10.2              | –      | 0.8      | –         | –             | –                   | –     | –                 | –      | –        | –        | –             |
| Portugal north | 2005              | 12,299| 7.1               | –      | 0.7      | –         | –             | –                   | –     | –                 | –      | –        | –        | –             |
| Republic of Ireland†† | 2005–2006 | 38,170| 5.6               | –      | 1.0      | 0.25     | 12.9          | 0.35                | –     | –                 | –      | –        | –        | –             |
| Spain        | 2005–2006          | 30,969| 4.5               | 0.8    | 0.5      | 0.15     | 6.8           | 0.47                | –     | –                 | –      | –        | –        | –             |
| Galicia††    | 2005–2006          | 30,969| 4.5               | 0.8    | 0.5      | 0.15     | 6.8           | 0.47                | –     | –                 | –      | –        | –        | –             |
| Spain Navarra| 2005–2006          | 734   | 13.9              | 2.2    | 1.4      | 0.41     | 6.9           | 0.43                | –     | –                 | –      | –        | –        | –             |
| Spain Valencia| 2005–2006         | 15,826| 12.6              | 1.7    | 1.0      | 0.26     | 5.7           | 0.36                | –     | –                 | –      | –        | –        | –             |
| Sweden Stockhol| 2005           | 8102  | 4.8               | –      | 0.7      | 0.17     | 10.2          | 0.35                | –     | –                 | –      | –        | –        | –             |
| Switzerland  | 2005                | 5790  | 7.4               | –      | 1.3      | 0.50     | 10.3          | 0.66                | –     | –                 | –      | –        | –        | –             |
| Fribourg     |                     |       |                   |        |          |          |                |                     |       |                   |        |          |          |               |
| The Netherlands | 2005              | 62,025| 2.7               | 1.2    | –       | –         | 19.9          | –                   | 668,238| 1.2              | 0.7    | –        | –        | 35.5          |
| UK England‡‡ | 2005–2007          | 531,870| 8.3              | 3.3    | 1.0      | 0.21     | 9.7           | 0.26                | 2,582,335| 3.7         | 1.5   | 0.9      | 0.07     | 21.6          |
| Total – all screening programmes | (2.2–15.6) | (2.2–15.6) | (0.8–3.3) | (0.04–1.4) | (0.06–0.50) | (4.9–24.2) | (0.18–0.66) | (2.2–15.6) | (1.2–10.5) | (0.3–1.5) | (0.3–1.1) | (0.04–0.10) | (6.8–49.5) | (0.07–0.21) |

Results are shown for screening programmes with <10% missing information on postoperative diagnosis.

1. 50–59 years old
2. Including initial tests, estimated to be <10% of total number of screening tests
3. 50–66 years old; screening test includes breast clinical examination
4. East Republic of Ireland, 50–64 years old
while the B/M ratio of surgical biopsy was 0.27 and 0.11, respectively, for initial and subsequent screening. There was less variation in the B/M ratio between programmes. False-positive rates can be calculated by subtraction of cancer detection rates from assessment rates. Therefore, in estimating the overall false-positive rates, we excluded those programmes not providing detection rates. We also excluded results for UK England, which were based on a three-year interval. After these exclusions, the overall false-positive rate was 5.4% in initial screens (6.0% further assessment rate, 0.55% detection rate) and 2.5% in subsequent screens (3.0% further assessment rate, 0.46% detection rate).

The European Union (EU) guidelines indicate that the desirable rate of further assessment should be <5% for initial and <3% for subsequent screens, but acceptable levels are <7% and <5%, respectively. The EUNICE data show that several (but not all) European programmes meet these targets. For initial screening, 42% (8/19) were acceptable and 26% (5/19) were desirable, while for subsequent screening 72% (13/18) were acceptable and 44% (8/18) desirable (Table 2).

The PPV of the screening test and the rates of needle biopsy and surgery (for diagnosis and treatment of screen-detected breast cancer) increased with increasing age, while the B/M ratio decreased with age (Table 3). In subsequent screening, the rate of further assessment was positively correlated with 1/PPV as might be expected, while no correlation was observed between the further assessment rate and needle biopsy, surgical intervention or surgical intervention with benign outcome (Figure 1). The rates of needle biopsy presented in Figure 1 and Table 2 are on average 24% and 28%, respectively, of the reported further assessment rates for initial and subsequent screens.

**DISCUSSION**

In mammography service screening, a biopsy is required to verify the presence of breast cancer, but it is important to keep the biopsy rate as low as possible in women not found to have cancer (false-positives), using the least invasive techniques. Inherent to a false-positive screening test is the psychological stress, assumed to be related to the fear of having breast cancer and the subsequent risk of death. However, the stress is mainly transient and related to the time between when the call is received and the suspicious finding is resolved. If a diagnosis of breast cancer is excluded with additional imaging and clinical examination only, the anxiety will usually begin to subside as soon as the outcome is communicated. Performance of a needle biopsy or a surgical biopsy and the subsequent waiting time for the result prolongs the stress, which might be related to an element of unpreparedness. Women with symptoms are probably more prepared to undergo further examinations than those who are recalled due to a positive screening mammogram. Fixed protocols for further assessment procedures communicated to the women and avoidance of short-term follow up (early re-screens), as recommended in the European guidelines, could reduce the psychological stress related to recalls. Specialist screening units with dedicated and well-trained staff (clinicians, administrative personnel, breast nurses and radiographers) equipped with good communication skills are more likely to achieve these aims than non-dedicated clinical practices.

Using data from three published observational studies, the estimated cumulative risk of having a recall with additional imaging but a benign result is 20% (ranging between 8% and 21%), and the risk of having an invasive procedure (i.e. needle biopsy or surgical intervention) is 3% (range 2–6%), among women aged 50–51 who attend 10 screening tests over two decades. Only one study reported the cumulative risk of a false-positive screening test resulting in surgical intervention (risk of 0.9%). The results were consistent with a fourth study based on seven screening tests. The pooled estimates should be interpreted with caution due to the small number of studies. The largest study has the longest follow-up, while the Norwegian study has the shortest. A substantial proportion of FNAC (3.9%) performed on cysts in the start-up period of the programme might explain the higher needle biopsy rate in Norway compared with the large study in Spain.

From the EUNICE data, the overall false-positive rate was 5.4% in initial screens (6.0% further assessment rate, 0.55% detection rate) and 2.5% in subsequent screens (3.0% further assessment rate, 0.46% detection rate), consistent

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Number of women with the actual information and screened</th>
<th>All further assessment [%]</th>
<th>Needle biopsy [%]</th>
<th>Surgical intervention (all) [%]</th>
<th>Surgical intervention with benign outcome [%]</th>
<th>PPV of the screening test [%]</th>
<th>B/M ratio of surgical biopsies</th>
</tr>
</thead>
<tbody>
<tr>
<td>50–54</td>
<td>1,152,188</td>
<td>4.8</td>
<td>0.89</td>
<td>0.52</td>
<td>0.09</td>
<td>10.9</td>
<td>0.22</td>
</tr>
<tr>
<td>55–59</td>
<td>1,972,213</td>
<td>3.8</td>
<td>1.00</td>
<td>0.63</td>
<td>0.07</td>
<td>17.1</td>
<td>0.12</td>
</tr>
<tr>
<td>60–64</td>
<td>1,577,939</td>
<td>3.7</td>
<td>1.18</td>
<td>0.78</td>
<td>0.07</td>
<td>21.3</td>
<td>0.10</td>
</tr>
<tr>
<td>65–69</td>
<td>1,238,820</td>
<td>4.0</td>
<td>1.50</td>
<td>0.99</td>
<td>0.07</td>
<td>24.4</td>
<td>0.08</td>
</tr>
<tr>
<td>Total</td>
<td>5,941,160</td>
<td>4.0</td>
<td>1.14</td>
<td>0.73</td>
<td>0.07</td>
<td>18.6</td>
<td>0.11</td>
</tr>
</tbody>
</table>

Poisson distribution **

P = 0.460

P = 0.031

P = 0.012

P = 0.490

P = 0.013

P = 0.070
with the results from the three observational studies\textsuperscript{10,23,24} and one research letter.\textsuperscript{25} Because of the three-year screening interval in the UK programme, England was excluded from this analysis along with programmes that did not provide information on detection rate. It should be noted that the observational studies are based on cohorts of women followed from age 50–51 on their first screen and attending successive screens, while the EUNICE survey represents a cross section of screens in a particular time period including participants with both regular and irregular attendance. The EUNICE data also show that the proportion of surgical intervention with benign outcome (Table 2) was well below the European desirable target of 0.25\textsuperscript{20} and declined compared with results in the early years of service screening.\textsuperscript{29} This target can probably be lowered in the next edition of the European Guidelines, due to improvements in imaging quality and a general shift from FNAC to larger CNB, which provides more diagnostic material.

The EUNICE data illustrate that the rates of further assessment and needle biopsy and surgery vary considerably between screening programmes in Europe. The personal characteristics of women screened might explain some of the substantial variation in further assessment rate and utilization of assessment procedures in European screening programmes. A recent study from Spain showed that the cumulative risk of a false-positive screening result varied from 8\% to 51\% in women with low-risk and high-risk profiles, respectively.\textsuperscript{18} Differences in other background conditions might also explain some of the observed variation.

The finding of a correlation between further assessment rate and PPV of the screening test in the EUNICE programmes (Figure 1) was expected.\textsuperscript{30} However, the lack of correlation between further assessment rate and needle biopsy, and with surgical intervention with benign histology was somewhat surprising, and means that a screening programme with high further assessment rates does not necessarily have high rates of invasive assessment. Figure 1 should be interpreted with caution given the relatively low number of observations, particularly for needle biopsy, for which information was provided by only nine out of the 18 countries. Additional analyses also showed that the \textit{P} value and rho did not differ substantially in Figures 1a, c and d when only the subset of 12 countries providing information on all relevant parameters was included in the analyses.
It should also be noted that, as expected, the assessment rate and the cancer detection rate were also positively associated (data available, not shown). The variability of detection rate among programmes was considerably lower than that of either the rate of further assessment or PPV, suggesting that it is possible to improve the detection rate without necessarily influencing the assessment rate.

An increase in PPV by age (Table 3) has been reported from other studies.29,31 This is reasonable, given the relatively higher breast density and lower incidence of disease at younger ages. Somewhat surprisingly, the further assessment rate did not decrease significantly by age, contrary to the findings of others.32 However, when we re-analysed excluding Ireland and Galicia, where women aged 65–69 were not invited, and England, which only recently initiated screening in this age group, the trend by age reached borderline significance (50–54: 5.4%; 55–59: 4.3%; 60–64: 4.0%; and 65–69: 3.8%, P = 0.090).

Continuous quality assurance and further improvements to optimize the screening process are needed to offer women in all European settings the quality achieved in programmes with the best results. Improvement in specificity over time has been demonstrated31 and may have an impact on future results.21 Analysis of the rate of further assessment should be performed in relation to the corresponding PPV of the screening test. It would arguably be preferable to have a recall rate of 4% and a PPV of 20%, implying a cancer detection rate of 0.8%, rather than a recall rate of 3% and a PPV of 15%, which would yield a cancer detection rate of 0.45%.

Key strengths of the EUNICE database include the large number of individuals for whom aggregate data are available, the large proportion of EU countries represented, the extensive feedback to project contributors, and data checking performed.21 There are several limitations to the present analysis of the EUNICE data. Given the aggregate format and relatively limited detail of the currently available data, it does not permit a precise estimation of cumulative false-positive rates. Differences in screening protocols, health-care systems and reporting systems in European countries may have affected the accuracy and completeness of data collection. An example of the way different screening protocols may limit the comparison and interpretation of data is that two out of the 20 programmes perform further assessment on the same day as screening, so that those with suspicious features on the mammograms undergo further investigation and receive a diagnosis on the same day as the screening.33 The expected avoidance of short-term anxiety associated with recall to assessment is accompanied, however, by an increase in further assessment rates. The reported correlations between the parameters calculated from the EUNICE data should be interpreted with care, as they are based on a relatively small number of observations (screening programmes).

Estimating the cumulative risk of a false-positive screening result is challenging due to a number of methodological issues, including the varying further assessment rates both within and between screening programmes. Different recall procedures may result from differences in screening regimens26,27 and in the characteristics of women attending screening (age, use of hormone treatment and previous breast biopsy).18,24 All the studies shown in Table 1 estimated the cumulative risk after 10 screening examinations, but none of the studies actually followed the women for 10 screening rounds. The cumulative risk is therefore based on estimated recall rates for four to seven screening rounds. This means that the risks of a false-positive test result might be overestimated in the studies cited in our review. An assumption of independence between screening results in subsequent screening rounds is discussed in several methodological studies, and different models for estimating the cumulative risk of false-positive screening results have been developed.34–36 The assumption of independence between recalls in subsequent screening rounds was confirmed in Hofvind et al.10 and Njor et al.23

In studies in the USA, the recall rate in a given round of screening has been reported to be as high as 15%15,37–39 and the cumulative risk of a false-positive screening result, with and without invasive procedures, is therefore considerably higher than the estimates in the European studies. Screening readers in most European screening programmes perform a specified volume of screening and diagnostic mammograms, in compliance with the EU guidelines.20 Adherence to the EU guideline recommendations is suggested as being the main reason for the lower recall rate, and therefore the lower cumulative risk of a false-positive screening test in European programmes compared with the results reported in the USA.10,23,24,36

**CONCLUSIONS**

Previously published estimates of the cumulative risk of a recall for further assessment among women aged 50–69 without breast cancer were reviewed. Over a period of two decades with biennial screening these varied from 8% to 21% in Europe (average 20%), and the cumulative risk of an invasive procedure was 3%. Performance monitoring data collected from several population-based service screening programmes in Europe are mostly within accepted levels according to the European guidelines, but outliers were identified. Continued quality assurance is required to offer women high quality mammographic screening.

**EUNICE WORKING GROUP**

The following authors were responsible for data collection from participating screening programmes or for data verification:
- Belgium: Van Hal G1, Martens P2; Czech Republic: Májek O1, Danes J1; Denmark: von Euler-Chelpin M1; Estonia: Aasmää A2; Finland: Aanttila A2; Germany: Becker N1, Hungary: Péntek Z1, Budai A2; Mádai S10; Ireland: Fitzpatrick P16, Mooney T16; Italy: Zappa M11, Ventura L11; Luxembourg: Scharpantgen A12; Norway: Hofvind S13; Poland: Seroczynski P14; Portugal: Morais A15, Rodrigues V15, Bento MJ15, Gomes de Carvalho J15; Spain: Natal C17, Prieto M17, Sánchez-Contador Escudero C18; Zubizarreta Alberti R19, Fernández Llanes SB19, Ascunce N20, Ederra Sanza M20, Sarriugarte Iriogoin G21; Salas Trejo D22; Ibañez Cabanell J22; Sweden: Wiege M23, Ohlsson G23, Törnberg S24; Korzeniewska M25; Switzerland: de Wolf C26; The Netherlands: Fracheboud
FALSE-POSITIVE MAMMOGRAPHY RESULTS: LITERATURE REVIEW

Members:

Ancelle-Park, R (F)1, Armaroli P (I)2, Ascunce N (E)3, Bisanti, L (I)4, Bellissario C (I)2, Broeders M (NL)5, Cogo C (I)6, De Koning H (NL)7, Duffy S W (UK)8, Frigerio A (I)7, Giordano L (I)2, Hofvind S (N)9, Jonsson H (S)10, Lyenge E (DK)11, Massat N (UK)6, Miccinesi G (I)12, Moss S (UK)8, Naldoni C (I)13, Njor S (DK)14, Nyström L (S)14, Paap E (NL)5, Paci E (I)12, Patnick J (UK)15, Ponti A (I)2, Puliti D (I)12, Segnan N (I)2, Von Karsa L (D)16, Törnberg S (S)17, Zappa M (I)12, Zorzi M (I)6

Affiliations:

1Ministère du travail et de la santé, Paris, France
2CPO-Piedmont, Turin, Italy
3Navarra Breast Cancer Screening Programme, Pamplona, Spain
4S.C. Epidemiologia, ASL di Milano, Italy
5Radboud University Nijmegen Medical Centre & National Expert and Training Centre for Breast Cancer Screening, Nijmegen, The Netherlands
6Veneto Tumor Registry, Padua, Italy
7Erasmus MC, Dept. Of Public Health Rotterdam, The Netherlands
8Wolfson Institute of Preventive Medicine, Queen Mary University of London, UK
9Cancer Registry of Norway, Research Department and Oslo and Akershus University College of Applied Science, Oslo, Norway
10Department of Radiation Sciences, Oncology, Umeå University, Umeå, Sweden
11Centre for Epidemiology and Screening, University of Copenhagen, Denmark
12ISPO Cancer Research and Prevention Institute, Florence, Italy
13Regional Cancer Screening Center, Emilia-Romagna Region, Bologna, Italy
14Department of Public Health and Clinical Medicine, Division of Epidemiology and Global Health, Umeå University, Umeå, Sweden
15NHS Cancer Screening Programmes and Oxford University, UK
16International Agency for Research on Cancer, Lyon, France
17Stockholm Cancer Screening, Stockholm, Sweden

Authors' affiliations

Solveig Hofvind, Researcher, Department of Research, Cancer Registry of Norway, Oslo, Norway
Antonio Ponti, MD MPH, Epidemiologist, Epidemiology Unit, CPO Piemonte, AOU S. Giovanni Battista, Turin, Italy
Julietta Patnick, Director, NHS Cancer Screening Programmes, Sheffield, UK
Nieves Ascunce, Public Health Doctor, Navarra Breast Cancer Screening Programme. Spanish Cancer Screening Network, Public Health Institute, Pamplona, Spain
Sisse Njor, Post Doc, Centre for Epidemiology and Screening, University of Copenhagen, Copenhagen, Denmark
Mireille Broeders, Senior Epidemiologist, Department of Epidemiology, Biostatistics and HTA, Radboud University Nijmegen Medical Centre, and National Expert and Training Centre for Breast Cancer Screening, Nijmegen, The Netherlands
Liviu Giordano, MD MPH, Epidemiologist, Epidemiology Unit, CPO Piemonte, AOU S. Giovanni Battista, Turin, Italy

EUROSCREEN WORKING GROUP

Coordinators:

E Paci, M Broeders, S Hofvind, S W Duffy

Department of Epidemiology and Biostatistics Department – National Institute for Health Development, Tallinn, Estonia
8German Cancer Research Center, Division of Cancer Epidemiology, Heidelberg, Germany
9National Mammography Screening Evaluation Committee (NMSEC), Budapest, Hungary
10MaMMa Healthcare Co., Budapest, Hungary
11National Centre of Screening Monitoring, Florence, Italy
12Programme Mammographic, Direzione de la Santé, Luxembourg
13The Cancer Registry of Norway, Oslo, Norway
14Asseco Poland S.A.
15Liga Portuguesa contra o Cancro, Lisbon, Portugal
16BreastCheck, National Cancer Screening Service, Dublin, Ireland
17Subdirección de Gestión Clínica y Calidad, Servicio de Salud del Principado de Asturias, Spain
18Dirección General de Salud Pública y Participación, Conselleria de Salut i Consum, Baleares, Spain
19Dirección General de Innovación y Gestión de la Salud Pública, Conselleria de Sanidade, Galicia, Spain
20Instituto de Salud Pública, Breast Cancer Screening Program, Navarra, Spain
21Programa Deteccio´n Precoz Ca´ncer de Mama, Direccio´n Territorial de Sanidad de Biza´kaia, Bilbao – Spain
22General Directorate Research and Public Health & Centre for Public Health Research (CSISP), Valencia, Spain
23Målsjukhuset, Eskilstuna, Sweden
24Cancer Screening Unit, Stockholm, Sweden
25Brosthenheten Våtärs Central Hospital, Västmanland, Sweden
26Swiss Federation of Breast Cancer Screening, Bern, Switzerland
27Department of Public Health, NETB, Erasmus University Medical Centre, Rotterdam, The Netherlands
28NHS Cancer Screening Programmes, Sheffield, UK
29Quality Assurance Group, Early Detection and Prevention Section, International Agency for Research on Cancer, Lyon, France
Alfonso Frigerio, Radiologist, Regional Reference Centre for Breast Cancer Screening, AO U. S. Giovanni Battista, Turin, Italy
Sven Törnberg, Oncologist and Director, Cancer Screening Unit, Oncologic Centre S3-00, Karlolinna University Hospital, Stockholm, Sweden

The EUNICE Project and Euroscreen Working Groups (Members of the EUNICE Project and Euroscreen Working Groups listed at end of paper)

ACKNOWLEDGEMENTS

The EUNICE breast cancer screening survey was financially supported by the European Commission (grant agreements no. 2004114, European Network for Information on Cancer, and no. 2006322, European Cancer Network). Financial support from the EU Public Health Programme (project no. 2006322, European Cooperation on development and implementation of Cancer screening and prevention Guidelines (ECCG-ECN)) was also provided for the preparation of this report. The contribution of Mariano Tomatis (CPO Piemonte) in designing the EUNICE data warehouse and in data management and analysis is gratefully acknowledged. Financial support was provided by the National Monitoring Italian Centre (ONS) to host the EUROSCREEN meetings in Florence in November 2010 and in March 2011 and the supplement publication and the National Expert and Training Centre for Breast Cancer Screening, Nijmegen, the Netherlands to host a meeting of the EUROSCREEN mortality working group in July 2011.

REFERENCES

5 van der Steeg AF, Keyzer-Dekker CM, De Vries J, Roukema JA. Effect of protocol-related variables on the risk of false-positive results in breast cancer screening programmes. J Epidemiol Community Health 2006;60:316–21