Overdiagnosing in breast cancer screening

Prof.dr.med. Per Skaane

Department of Radiology, Oslo University Hospital Ullevaal, University of Oslo, Oslo, Norway

PERSKA@ous-hf.no
Objectives: Overdiagnosing in breast cancer screening

- Definition: What is overdiagnosis?
- If overdiagnosis is a challenge for radiologists: Why?
- The «never-ending» Scandinavian quarrel
- What should we believe regarding overdiagnosis and Conclusion
Adverse effects of screening mammography

A) Definition: What is overdiagnosis?

- Mammographic examination
  - Discomfort (painful)
  - Mobilizing cancer cells?
  - Radiation dose
- False positive result (recalls)
- False negative result (interval cancer)
- Overdiagnosis ("overdetection")
- Overtreatment
Overdiagnosis:
The diagnosis of a histologically proven cancer that would never have been diagnosed in the woman’s lifetime in the absence of mammography screening.
Overdiagnosis («what are we talking about»?):

Overdiagnosis* in mammographic screening is the diagnosis of (low grade 1) invasive cancers (?) and/or DCIS that, in the absence of screening, would not have presented clinically during the woman’s lifetime

* Overdiagnosis: An extreme form of length bias!
Theory of linear progression


Theory of parallel disease

Wiechmann L: Cancer 2008;112:2130
Growth rates of breast cancers

- Fast
- Slow
- Very slow
- Non-progressive

Clinically evident
- Screening rounds
- Time
- Mortality

Tumor size
Screening bias

A) Lead-time bias: Earlier detection but no impact on date of death
B) Length-time bias: Indolent tumors more likely to be detected

Overdiagnosis: Extreme form of length-time bias
Women aged 50-74 years: Breast cancer incidence rates

<table>
<thead>
<tr>
<th>Incidence rates (x 1000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted</td>
</tr>
<tr>
<td>Observed</td>
</tr>
<tr>
<td>Observed corrected for lead time</td>
</tr>
</tbody>
</table>

Excess ratio for invasive cancer and DCIS:
• In total, observed to predicted cases = 36.2 %
• After correction for lead time = 4.6 %

Incidence of Ductal Carcinoma *in situ* in the period before, during and after implementation of a population-based screening program

Sørum R¹, Hofvind S¹/², Skaane P³ and Haldorsen T¹

¹The Cancer Registry of Norway, Department of screening-based research, Oslo, Norway. ²Oslo University College, Faculty of Health Sciences, Oslo, Norway. ³Ullevaal University Hospital, Department of Radiology, University of Oslo, Oslo, Norway

*Figure 1*: Age-adjusted incidence of primary pure DCIS in Norway, 1993-2007
FIGURE 16.5 Expected breast cancer incidence with screening (green line) and without screening (red line). Wishful thinking, as the huge drop in elderly women has never been observed in practice (redrawn).

Zahl PH, Mæhlen J: Overdiagnostikk av brystkreft etter 14 år med mammografiscreening
Tidsskr Nor Legeforen 2012;132:414-417

**Figur 1** Alderspesifikk insidens av invasiv brystkreft i prøvefylkene i perioden 1998–2009 (rød linje) og i perioden 1991–95 (gul linje). Merk at de to første aldersgruppene er femårige (40–44 år og 45–49 år) og at aldersgruppene er toårlige fra og med 50 år.
Overdiagnosis among women attending a population-based mammography screening program

B) If overdiagnosis is a challenge for radiologists: Why?

Overdiagnosis likely to be driven by (??):

• The radiologist’s fear of missing a cancer, and the fear of litigation

• Technological developments:
  - Digital mammography (FFDM)
  - Computer aided detection (CAD)
  - Improved biopsy techniques (VAB)
  - Tomosynthesis (DBT)
### Studies comparing SFM and FFDM in breast cancer screening: DCIS detection rate and proportion of DCIS in FFDM*

<table>
<thead>
<tr>
<th>Study</th>
<th>Examinations (n)</th>
<th>DCIS detection Rate (%)</th>
<th>Proportion DCIS (FFDM)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SFM</td>
<td>FFDM</td>
<td>SFM</td>
</tr>
<tr>
<td>Oslo II</td>
<td>16,985</td>
<td>6,944</td>
<td>0.12</td>
</tr>
<tr>
<td>DMIST a)</td>
<td>42,760</td>
<td>42,760</td>
<td>0.12</td>
</tr>
<tr>
<td>Florence b)</td>
<td>14,385</td>
<td>14,385</td>
<td>0.12</td>
</tr>
<tr>
<td>Vestfold c)</td>
<td>324,763</td>
<td>18,239</td>
<td>0.11</td>
</tr>
<tr>
<td>INBSP*</td>
<td>153,619</td>
<td>35,204</td>
<td>0.09</td>
</tr>
<tr>
<td>DSPP* d)</td>
<td>311,082</td>
<td>56,518</td>
<td>0.08</td>
</tr>
</tbody>
</table>

* Studies with no specification of DCIS or a small number of DCIS excluded
a) Cancers diagnosed within 455 days after imaging
b) Cancers presenting as clustered microcalcifications
c) Prevalent screening rounds; SFM is mean value of merged data from 18 counties
d) CAD used for FFDM only

*) INBSP: Irish National Breast Screening Program
*) DSPP: Digital Screening Project Preventicon
Oslo I Study FFDM: Independent double-reading 1 - 1

Cancer in subsequent round:
IDC and DCIS Grade 3
Diagnosis of DCIS

BI-RADS – Mammography: Microcalcifications*

Morphologic descriptors

1. Typically benign
2. Suspicious (intermediate concern)
   - Amorphous (indistinct) calcifications
   - Coarse heterogeneous calcifications
3. Higher probability of malignancy
   - Fine pleomorphic calcifications
   - Fine linear / fine linear branching

Distribution descriptors

Stability descriptors**

* American College of Radiology: Breast Imaging Reporting and Data System
** not officially included in the BI-RADS lexicon
Microcalcifications: The great challenge!

US - guided FNAC: DCIS high grade (G3)
Improved biopsy techniques:
Vacuum Assisted Biopsy (VAB)

Histology: DCIS grade 3, 10 mm
Tomosynthesis: Potential for increased sensitivity

Invasive lobular carcinoma (ILC): 2 mm, gr. 1
(+ LCIS 20 mm and ALH / LCIS)
DCIS gr. 1:

- Mammographic extent ("main group") 10 mm
- Histology extent (minimum) 35 mm
B) Interpretation error:

Amorphous and indistinct/heterogeneous calcifications:
Linear distribution elevates suspicion of any morphology!!

Histology (mastectomy): DCIS gr. 3; extent 50 mm (with comedo-necroses and multiple foci of microinvasive cancer less than 1 mm)
Interpretation error:

Oslo Tomosynthesis Screening Trial

<table>
<thead>
<tr>
<th>Radiologist</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score (2010)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>
Interpretation error:

Disappearance of calcifications: Development of invasive cancer!!
( note also development of casting-type calcification )

Histology: DCIS gr. 3 + IDC
<table>
<thead>
<tr>
<th>Grade</th>
<th>Size (mm)</th>
<th>Casting no. (%</th>
<th>Granular no. (%)</th>
<th>Punctate no. (%)</th>
<th>Total no.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 10</td>
<td>253 (50)</td>
<td>203 (40)</td>
<td>55 (11)</td>
<td>511</td>
<td></td>
</tr>
<tr>
<td>10-20</td>
<td>223 (59)</td>
<td>120 (31)</td>
<td>38 (10)</td>
<td>381</td>
<td></td>
</tr>
<tr>
<td>21-30</td>
<td>87 (76)</td>
<td>20 (18)</td>
<td>7 (6)</td>
<td>114</td>
<td></td>
</tr>
<tr>
<td>&gt; 30</td>
<td>94 (77)</td>
<td>22 (18)</td>
<td>6 (5)</td>
<td>122</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>657 (58)</td>
<td>365 (32)</td>
<td>106 (9)</td>
<td>1128</td>
<td></td>
</tr>
<tr>
<td><strong>Intermed</strong></td>
<td>Total</td>
<td>182 (37)</td>
<td>208 (43)</td>
<td>95 (20)</td>
<td>485</td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 10</td>
<td>22 (22)</td>
<td>55 (54)</td>
<td>24 (24)</td>
<td>101</td>
<td></td>
</tr>
<tr>
<td>10-20</td>
<td>11 (26)</td>
<td>20 (48)</td>
<td>11 (26)</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>21-30</td>
<td>6 (43)</td>
<td>6 (43)</td>
<td>2 (14)</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>&gt; 30</td>
<td>6 (46)</td>
<td>6 (46)</td>
<td>1 (8)</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>45 (26)</td>
<td>87 (51)</td>
<td>38 (22)</td>
<td>170</td>
<td></td>
</tr>
</tbody>
</table>

* Evans A et al.: The Sloane project (a large, multicentre audit of screen-detected DCIS in the UK). Clin Radiol 2010;65:181-184 (modified)
C) The «never-ending» Scandinavian quarrel
Overdiagnosis of Invasive Breast Cancer due to Mammography Screening

Per-Henrik Zahl, DrMedSci

Kalager and colleagues (1) reported a 15% to 25% overdiagnosis rate of invasive cancer in the Norwegian Breast Cancer Screening Program. In 2004, my coworkers and I reported (2) that screening led to a 51% increase in breast cancer cases that could not be explained by earlier diagnosis and called this increase overdiagnosis. We updated our study in 2012 (3) and reported the same high level of overdiagnosis. Our updated study had longer follow-up and included 11,631 breast cancer cases diagnosed after screening started, whereas Kalager and colleagues’ analysis included only 3,862 cases. The difference in estimates of overdiagnosis between our study and theirs, however, is not explained by data per se but by the methods for adjusting for underlying incidence increase and lead time.
Overdiagnosis in mammography screening.

Per-Henrik Zahl, Jan Mæhlen

Background. In Norway and Sweden, the introduction of mammography screening programmes has been associated with about a 50% increase in breast cancer incidence for the screened age groups and almost stable incidence in higher age groups. This suggests that mammography screening results in a substantial degree of overdiagnosis.

EJSO (2004) 30, 711-712

EDITORIAL

Overdiagnosis of breast cancer in screening

The recent paper by Zahl et al.1 has again raised doubts about the supposed benefits of breast screening as a public health policy. Theoretically, the introduction of breast screening in risk population should (and does) detect more cancers specifically because of the lead and length time bias, which make the evaluation of a screening programme so difficult. A consequence of these factors is overdiagnosis of breast cancer which is defined as the detection of low malignancy lesions that would otherwise have not been detected in a patients lifetime.

The recent reported rates of ‘overdiagnosis’ of breast cancer in Norway (54%) and Sweden (45%) appear to be higher than those expected2 or reported previously4 in clinical trials. However, another recent study of increase in incidence of The detractors of breast screening point out that many of these tumours would never become symptomatic, and although it has been shown in several studies that breast cancer specific mortality has decreased in countries which have implemented a national screening programme,11-13 overall mortality has in general not followed this trend.14 Many would argue that overall mortality should not be the benchmark for the evaluation of a breast screening programme in various countries has coincided with better surgery and with the introduction of effective adjuvant treatment for breast cancer, the reductions in mortality from breast cancer are not

Zahl PH al.: Tidsskr Nor Lægeforen 2007;127:207-8
Norway

Breast cancer per year: about 2700

This is one of the greatest scandals in modern medicine
Legg ned mammografiprogrammet

Mammografiprogrammet best uten mammografi?

KRONIKK

Nils August Andresen
Redaktør i Minerva

MEDI NY TEKNOLOGI kan vi behandle stadig flere for stadig mer. At dette koster stadig mer

HELESE: Hvis å være føde var betyr å påføre sykdom, smerte, angst og lidelse på kvinner og familier, da er jeg imot.

Mammografiprogrammet

Legg ned mammografiprogrammet, skrev Nils August Andresen i Dagbladet 30. mai. Dette står i sterk motsetning til hjemmesidene til si at den er reell (såkalt statistisk signifikans). Nedgangen ligger langt under de 30 prosent som Stortinget forventet da programmet startet i 1996.

Resultatene av studien er for øvrig akkurat like som i en annen stor studie, som i 2010 ble publisert i det mest anerkjente medisinske tidskriften New England Journal of Medicine og som viste en nedgang på ti pro-

programmet ikke hjelper, er det nemlig slik, sies det, at mange kvinner allerede gikk til mammografiscreening før programmet begynte. Derfor kunne man ikke finne en nedgang etter at programmet ble innført. Man snur resultatene på hodet og sier at mammografiprogrammet «egenlig» virker, selv om tallene sier det motsatte.
Increased incidence of ductal carcinoma in situ (DCIS) and invasive breast cancer (IBC) after introduction of organized screening has prompted debate about overdiagnosis. The aim was to examine the excess in incidence of DCIS and IBC during the screening period and the deficit after women left the program, and thereby to estimate the proportion of overdiagnosis. Women invited to the Norwegian Breast Cancer Screening Program were analyzed for DCIS or IBC during the period 1995–2009. Incidence rate ratios (IRRs) were calculated for attended vs. never attended women. The IRRs were adjusted by Mantel-Haenszel (MH) method and applied to a set of reference rates and a reference population to estimate the proportion of overdiagnosis during the women's lifespan after the age of 50 years. A total of 702,131 women were invited to the program. An excess of DCIS and IBC was observed among women who attended screening during the screening period; prevalently invited women aged 50–51 years had a MH IRR of 1.86 (95% CI 1.65–2.09) and subsequently invited women aged 52–69 years had a MH IRR of 1.56 (95% CI 1.45–1.68). In women aged 70–79 years, a deficit of 30% (MH IRR 0.70, 95% CI 0.62–0.80) was observed 1–10 years after they left the screening program. The estimated proportion of overdiagnosis varied from 10 to 20% depending on outcome and whether the women were invited or actually screened. The results highlight the need for individual data with longitudinal screening history and long-term follow-up as a basis for estimating overdiagnosis.
Overdiagnosis among women attending a population-based mammography screening program


Table 2. Estimated proportion of overdiagnosis among attending and invited women after implementation of the screening program

<table>
<thead>
<tr>
<th>Reference</th>
<th>Attended(^1)</th>
<th>Invited(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OD (%)</td>
<td>95% CI</td>
</tr>
<tr>
<td>DCIS and invasive breast cancer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modeling approach</td>
<td>19.6</td>
<td>12.1–27.1</td>
</tr>
<tr>
<td>Period approach</td>
<td>19.4</td>
<td>11.8–27.0</td>
</tr>
<tr>
<td>Cohort approach</td>
<td>16.5</td>
<td>9.1–23.9</td>
</tr>
<tr>
<td>Invasive breast cancer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modeling approach</td>
<td>13.4</td>
<td>4.7–22.1</td>
</tr>
<tr>
<td>Period approach</td>
<td>13.3</td>
<td>4.0–22.6</td>
</tr>
<tr>
<td>Cohort approach</td>
<td>11.4</td>
<td>2.7–20.1</td>
</tr>
</tbody>
</table>

\(^1\)Women who attended screening compared with women who never attended.

\(^2\)Calculated as the estimate among attending women multiplied by the compliance of the program (84%).

Abbreviations: DCIS: ductal carcinoma in situ; OD: estimated proportion of overdiagnosis; CI: confidence interval.
Natural history of breast cancer and screening

No detectable disease

Disease detectable by screening

Clinical Symptomatic disease

Lead time

Sojourn time

Screening

Time
Overdiagnosis in breast cancer screening: the importance of length of observation period and lead time

Stephen W Duffy* and Dharmisha Parmar

Abstract

Background: Overdiagnosis in breast cancer screening is a controversial topic. One difficulty in estimation of overdiagnosis is the separation of overdiagnosis from lead time that is the advance in time of diagnosis of cancers, which confers an artificial increase in incidence when a screening programme is introduced.

Methods: We postulated a female population aged 50-79 with a similar age structure and age-specific breast cancer incidence as in England and Wales before the screening programme. We then imposed a two-yearly screening programme; screening women aged 50-69, to run for twenty years, with exponentially distributed lead time with an average of 40 months in screen-detected cancers. We imposed no effect of the screening on incidence other than lead time.

Results: Comparison of age- and time-specific incidence between the screened and unscreened populations showed a major effect of lead time, which could only be adjusted for by follow-up for more than two decades and including ten years after the last screen. From lead time alone, twenty-year observation at ages 50-69 would confer an observed excess incidence of 37%. The excess would only fall below 10% with 25 years or more follow-up. For the excess to be nullified, we would require 30 year follow-up including observation up to 10 years above the upper age limit for screening.

Conclusion: Studies using shorter observation periods will overestimate overdiagnosis by inclusion of cancers diagnosed early due to lead time among the nominally overdiagnosed tumours.

Keywords: screening; lead time; overdiagnosis

Overestimated lead times in cancer screening has led to substantial underestimation of overdiagnosis

P-H Zahl*,1, K J Jørgensen2 and P C Gøtzsche2

1Norwegian Institute of Public Health, PO Box 4404 Nydalen, N-0403 Oslo, Norway and 2The Nordic Cochrane Centre, Rigs hospitalet, Blegdamsvej 9, DK-2100 Copenhagen, Denmark
## Overdiagnosis in breast cancer screening

<table>
<thead>
<tr>
<th>Author</th>
<th>Publ.</th>
<th>Study</th>
<th>Journal</th>
<th>Overdiagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paci E</td>
<td>2004</td>
<td>Florence</td>
<td>J Med Screen</td>
<td>&lt; 5 %</td>
</tr>
<tr>
<td>Zahl PH*</td>
<td>2004</td>
<td>Norway Sweden</td>
<td>BMJ</td>
<td>54% 45%</td>
</tr>
<tr>
<td>Duffy SW</td>
<td>2005</td>
<td>Two-county Gothenburg</td>
<td>Breast Cancer Research</td>
<td>1 %</td>
</tr>
<tr>
<td>de Koning</td>
<td>2006</td>
<td>Netherlands</td>
<td>Breast Cancer Research</td>
<td>3 %</td>
</tr>
<tr>
<td>Zackrisson S</td>
<td>2006</td>
<td>Malmø</td>
<td>BMJ</td>
<td>10 %</td>
</tr>
<tr>
<td>Olsen AH</td>
<td>2006</td>
<td>Copenhagen</td>
<td>Breast Journal</td>
<td>5 %</td>
</tr>
<tr>
<td>Gøtzsche PC</td>
<td>2006</td>
<td>Overview</td>
<td>Cochrane Library</td>
<td>30 %</td>
</tr>
</tbody>
</table>

de Koning et al.: Breast Cancer Research 2006; 8: 202  
Zackrisson S et al.: BMJ 2006; 332: 689-691  
Results  There were 13 primary studies reporting 16 estimates of overdiagnosis in seven European countries (the Netherlands, Italy, Norway, Sweden, Denmark, UK and Spain). Unadjusted estimates ranged from 0% to 54%. Reported estimates adjusted for breast cancer risk and lead time were 2.8% in the Netherlands, 4.6% and 1.0% in Italy, 7.0% in Denmark and 10% and 3.3% in England and Wales.

Conclusions  The most plausible estimates of overdiagnosis range from 1% to 10%. Substantially higher estimates of overdiagnosis reported in the literature are due to the lack of adjustment for breast cancer risk and/or lead time.
Evaluation of the Norwegian Breast Cancer Screening Program (NBCSP):

Kalager M et al. NEJM 2010:
- Aggregated data
- Short follow-up

Mortality reduction: 10%

Hofvind S et al. Cancer 2013:
- Individual data
- Long-term follow-up

Mortality reduction: 43%
There is no such thing as overdiagnosis: there is only correct, partially correct, or incorrect diagnosis. If abnormal findings are diagnosed correctly, there is only optimally managed, suboptimally managed, mismanaged, and possibly overtreated disease.
Information to the women regarding «overdiagnosis»
What should we tell them ????

Attendance rate in Norwegian Breast Cancer Screening Program 2005-2012: All counties

<table>
<thead>
<tr>
<th>Year</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate</td>
<td>76%</td>
<td>77%</td>
<td>75%</td>
<td>75%</td>
<td>75%</td>
<td>74%</td>
<td>75%</td>
<td>75%</td>
</tr>
</tbody>
</table>

Thank you very much for your time!

PERSKA@ous-hf.no

Harald Sohlberg (1869-1935):
Winter night in Rondane (Norway) (1914)