Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review

Donella Puliti, Stephen W Duffy, Guido Miccinesi, Harry de Koning, Elsebeth Lynge, Marco Zappa, Eugenio Paci and the EUROSSCREEN Working Group (members listed at the end of the paper)

**Objectives** Overdiagnosis, the detection through screening of a breast cancer that would never have been identified in the lifetime of the woman, is an adverse outcome of screening. We aimed to determine an estimate range for overdiagnosis of breast cancer in European mammographic service screening programmes.

**Methods** We conducted a literature review of observational studies that provided estimates of breast cancer overdiagnosis in European population-based mammographic screening programmes. Studies were classified according to the presence and the type of adjustment for breast cancer risk [data, model and covariates used], and for lead time (statistical adjustment or compensatory drop). We expressed estimates of overdiagnosis from each study as a percentage of the expected incidence in the absence of screening, even if the variability in the age range of the denominator could not be removed. Estimates including carcinoma in situ were considered when available.

**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.
seting over 20 years ago, before the implementation of service screening. It should also be noted that in the Canadian studies there remained considerable expected life years in which the control group might catch up further. It is therefore important to estimate overdiagnosis in the service screening setting in order to understand how technological advances and developments in the practice of screening have modified the risk of overdiagnosis.

We conducted a review of the European observational studies evaluating overdiagnosis of breast cancer in mammographic service screening programmes.

METHODS
Primary research articles that gave explicit estimates of breast cancer overdiagnosis in European population-based mammographic screening programmes, published in English, were eligible for inclusion in this review. Estimates including carcinoma in situ were considered when available.

The search strategy is provided in the Appendix; 133 English language abstracts pertinent to the review were considered. We excluded 36 editorials or commentaries, 22 reviews, 14 letters and 44 papers because they did not report an original estimate of overdiagnosis, one paper because it pertained to a non-European country and four papers reporting only results from randomized trials. On the basis of the references in the articles identified, one more paper was also included. We replaced one paper with an updated report from the same population using the same methodology and published after our search date. The list of the 13 selected studies8–20, classified according to the characteristics of the Population, Intervention, Comparison and Outcomes of each paper (PICO frame), is presented in Table 1.

For each selected paper, we defined the population by specifying three characteristics:

1. The country to which it referred;
2. The period of the study, defined as incidence calendar years – pre- and postscreening – included in the analysis;
3. The type of population. Population types were either demographic (i.e. a dynamic population analysing temporal trends and/or geographical differences), or cohort (if a defined population of subjects was followed up prospectively). Cohorts were further divided into two types, birth cohort and cohort by enrolment.

Adjustment for breast cancer risk and correction for lead time bias
In our review, we took into account some methodological issues and potential sources of bias in overdiagnosis estimation. Overdiagnosis can be correctly estimated by comparing incidence in screened and unscreened populations provided that (i) there are similar underlying risks of breast cancer in the two populations, and (ii) the effect of lead time (the period of time by which the diagnosis is brought forward by screening) is accounted for.

Adjustment for underlying breast cancer risk
A valid comparison group (the so-called ‘unscreened population’) should include women with comparable age span and with an underlying risk of breast cancer similar to the screened population. Adjustment for differences in the underlying risk between the screened and unscreened populations should be based on known risk factors for breast cancer (such as age, use of hormone replacement therapy, obesity, fertility rate, etc.). When the incidence in the unscreened population is derived from the prescreening period, an adjustment for the temporal trend in breast cancer risk is needed. When the incidence in the unscreened population is derived from a contemporaneous location in which there was no screening, an adjustment for prescreening geographical differences is required.

Adjustment for lead time
The major difficulty in the estimation of overdiagnosis is disentangling the excess of incidence due to lead time from the excess due to overdiagnosis. The excess incidence due to lead time (i.e. the increase in incidence after screening starts) is an expected and necessary outcome of breast cancer screening, reflecting the detection of cancers at a more treatable stage by bringing the diagnosis forward. The initial increase in breast cancer incidence in the screened group will persist while the women continue to be screened, because of the shift in the age–incidence curve. After the end of screening, a reduction of the incidence in the screened group should occur due to the earlier diagnosis of cancers in the screening period.

In the absence of overdiagnosis, the initial increase in breast cancer occurrence in the screened group would be fully compensated for by a similar decrease in cancers among older age groups no longer offered screening, the so-called ‘compensatory drop’. The compensatory drop method requires that the screening programme has been running long enough to achieve a full adjustment for lead time, i.e. substantial numbers of women should have actually been through the screening programme, have gone beyond the upper age limit and have a sufficient follow-up after screening stops (at least 5 years on the basis of the estimate of the breast cancer mean sojourn time5,13). Even with long-term observation, the compensatory drop method will slightly overestimate overdiagnosis, unless every screened cohort is followed up long past the upper age limit. For example, if screening were offered to women aged 50–69 and there were data to 2003, the women screened at age 65–69 in 2000–2003 will have a lead time excess, the compensatory drop of which would not be observable until after our period of observation.

If there is short or no follow-up after the last screen, there will be a lead time bias that should be adjusted for using statistical methods. The so-called ‘postponement of screen-detected cases’ method is used in some studies, wherein the dates of diagnosis of screen-detected cases are postponed for a period corresponding to the estimated lead time in order to calculate the incidence corrected for lead time.

In this review, we distinguish studies that used a compensatory drop method from those that used a statistical adjustment for lead time.
<table>
<thead>
<tr>
<th>Paper</th>
<th>Country</th>
<th>Calendar period</th>
<th>Type of population</th>
<th>Age and interval of screening</th>
<th>Start year of screening</th>
<th>Reference population</th>
<th>Adjustment for breast cancer risk</th>
<th>Adjustment for lead time</th>
<th>Measure of overdiagnosis (OD)†</th>
<th>Estimate OD (only invasive)</th>
<th>Estimate OD (in situ and invasive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peeters et al. (1989)9</td>
<td>The Netherlands (Nijmegen)</td>
<td>1970–1986</td>
<td>Dynamic population</td>
<td>35+ years biennial</td>
<td>1975</td>
<td>Incidence of neighbouring unscreened area</td>
<td>Birth year</td>
<td>No adjustment</td>
<td>Excess incidence/Expected incidence</td>
<td>Not reported</td>
<td>11%</td>
</tr>
<tr>
<td>Paci et al. (2004)9</td>
<td>Italy (Florence)</td>
<td>1985–1999</td>
<td>Dynamic population</td>
<td>50–69 years biennial</td>
<td>1990</td>
<td>Prescreening incidence</td>
<td>Age</td>
<td>Statistical adjustment</td>
<td>Adjusted excess incidence/Expected incidence</td>
<td>1%</td>
<td>5%</td>
</tr>
<tr>
<td>Zahl et al. (2004)10</td>
<td>Norway (AORH counties)</td>
<td>1971–2000</td>
<td>Dynamic population</td>
<td>50–69 years biennial</td>
<td>1996</td>
<td>Prescreening incidence</td>
<td>Age</td>
<td>No adjustment</td>
<td>Excess incidence/Expected incidence</td>
<td>54%</td>
<td>Not reported</td>
</tr>
<tr>
<td>Jonsson et al. (2005)11</td>
<td>Sweden (11 counties)</td>
<td>1971–2000</td>
<td>Dynamic population</td>
<td>Different age range: the broadest 40–74 the less extensive 50–69</td>
<td>1986–1990</td>
<td>Prescreening incidence</td>
<td>Age, temporal trend and area</td>
<td>Statistical adjustment</td>
<td>Adjusted excess incidence/Expected incidence</td>
<td>0–54% according to age</td>
<td>Not reported</td>
</tr>
<tr>
<td>Olsen et al. (2006)12</td>
<td>Denmark (Copenhagen)</td>
<td>1991–1996</td>
<td>Cohort by enrolment</td>
<td>50–69 years biennial</td>
<td>1991</td>
<td>Incidence among screened women</td>
<td>Not necessary</td>
<td>Statistical adjustment</td>
<td>Adjusted excess incidence/Observed incidence</td>
<td>Not reported</td>
<td>4.8%</td>
</tr>
<tr>
<td>Paci et al. (2006)13</td>
<td>Italy (Northern and Central areas)</td>
<td>1986–2001</td>
<td>Dynamic population</td>
<td>50–69 years biennial</td>
<td>1991–1997</td>
<td>Prescreening incidence</td>
<td>Age, temporal trend and area</td>
<td>Statistical adjustment</td>
<td>Adjusted excess incidence/Expected incidence</td>
<td>3.2%</td>
<td>4.6%</td>
</tr>
<tr>
<td>Waller et al. (2007)14</td>
<td>England and Wales</td>
<td>1971–2001</td>
<td>Dynamic population</td>
<td>50–64 years triennial (extended to 65–70)</td>
<td>1988 (2002)</td>
<td>Pre- and postscreening incidence</td>
<td>Age, period, birth cohort and use of HRT</td>
<td>Compensatory drop</td>
<td>Absolute increase of lifetime risk of breast cancer</td>
<td>0.86%</td>
<td>Not reported</td>
</tr>
</tbody>
</table>
### Table 1: Characteristics of studies included in the review

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Year of cohort</th>
<th>Age at screening</th>
<th>Duration</th>
<th>Data source</th>
<th>Incidence measure</th>
<th>Overdiagnosis measure</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Puliti et al. (2009)</td>
<td>Italy (Florence)</td>
<td>1968–2004</td>
<td>50–69 years</td>
<td>Biennial</td>
<td>Prescreening</td>
<td>Age and temporal trend</td>
<td>Compensatory drop</td>
<td>0%</td>
</tr>
<tr>
<td>Jorgensen et al. (2009)</td>
<td>Denmark (Funen and Copenhagen)</td>
<td>1971–2003</td>
<td>50–69 years</td>
<td>Biennial</td>
<td>Incidence of neighbouring unscreened area</td>
<td>Age and geographical area</td>
<td>Compensatory drop</td>
<td>Not reported</td>
</tr>
</tbody>
</table>
| Martinez-Alonso et al. | Spain (Catalonia) | 1980–2004      | 50–64 years      | Biennial | Pre- and post-screening incidence | Age, year of birth, fertility rate and use of mammography | Statistical adjustment | 2.3%
| de Gelder et al. (2011)| The Netherlands | 1989–2006      | 50–69 years      | Biennial | Predicted incidence without screening by MISCAN | Not necessary | Compensatory drop | Not reported |

*The year within brackets refers to the start of screening for the age cohort reported in brackets in the previous column. The range 1991–1997 for the paper by Paci et al. indicates the range of beginning of screening among areas included in the study. The range 1991–1993 for the paper by Jorgensen et al. indicates that screening started in 1991 in Copenhagen and in 1993 in Funen.

**Adj*’ed’ means ‘adjusted for lead time’.

†The compensatory drop was observed by authors (11% in Norway and 12% in Sweden) but it was not considered to estimate overdiagnosis because it was not statistically significant.

---

**Results**

We included 13 primary studies in our review, reporting 16 measures of overdiagnosis from population-based mammographic screening in seven Western European countries (the Netherlands, Italy, Norway, Sweden, Denmark, UK, and Spain). Table 1 shows the PICO description of 13 selected papers.

Table 1 indicates that the number of cases detected in each study was evaluated. We classified the papers by adjustment for lead time bias and breast cancer risk. However, the variability in the age range to which the data apply and the range is therefore reported in the table.
<table>
<thead>
<tr>
<th>Paper</th>
<th>Model</th>
<th>Adjusted for...</th>
<th>Comments</th>
<th>Annual increase in the prescreening period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peeters et al. (1989)</td>
<td>Mantel-Haenszel</td>
<td>Birth year</td>
<td>Birth-cohort specific incidence rates in a neighbouring unscreened area were used as reference</td>
<td>Not estimated</td>
</tr>
<tr>
<td>Paci et al. (2004)</td>
<td>Standardization</td>
<td>Age class (5-years)</td>
<td>The annual percentage change was reported in the tables of the Results but it was not taken into account in estimating overdiagnosis</td>
<td>Not estimated</td>
</tr>
<tr>
<td>Zahl et al. (2004)</td>
<td>Poisson regression over all the study period</td>
<td>Age class (50–69, 70–74)</td>
<td>The annual percentage change was reported in the tables of the Results but it was not taken into account in estimating overdiagnosis</td>
<td>1.01 (0.99–1.02)</td>
</tr>
<tr>
<td>Zahl et al. (2004)</td>
<td>Poisson regression over all the study period</td>
<td>Age class (50–69, 70–74, 75–79)</td>
<td>The annual percentage change was reported in the tables of the Results but it was not taken into account in estimating overdiagnosis</td>
<td>1.008 (1.007–1.009)</td>
</tr>
<tr>
<td>Jonsson et al. (2005)</td>
<td>Poisson regression on prescreening incidence (15 years of prescreening)</td>
<td>Year and county, stratified for four age class (40–49, 50–59, 60–69, 70–74)</td>
<td>The model included the year on the screening scale (from −15 to 1) instead of the calendar year</td>
<td>Not reported</td>
</tr>
<tr>
<td>Olsen et al. (2006)</td>
<td>Multistate modelling</td>
<td>Not necessary</td>
<td>The estimate of overdiagnosis was obtained fitting a model to the screening data. No adjustment for breast cancer risk is needed</td>
<td>Not estimated</td>
</tr>
<tr>
<td>Paci et al. (2006)</td>
<td>Poisson regression on prescreening incidence (6 years of prescreening)</td>
<td>Age (annual, 40–79), calendar year and area</td>
<td>Two-step Poisson analysis: 1) A first model (age, year) was fitted to the available prescreening incidence for each area 2) A pooled model (age, year, area) was fitted to prescreening incidence observed or estimated by model 1 (wherever missing)</td>
<td>1.012 (1.008–1.016)</td>
</tr>
<tr>
<td>Waller et al. (2007)</td>
<td>Age-period-cohort model over all the study period</td>
<td>Age (annual, 20–89), calendar year, birth cohort, use of HRT and screening attendance</td>
<td>Estimates of the prevalence of HRT use were obtained from the General Practice Research Database</td>
<td>Not reported</td>
</tr>
<tr>
<td>Jørgensen and Gotzsche (2009)</td>
<td>Linear regression on prescreening incidence (1971–84)</td>
<td>Calendar year, stratified for three age class (30–49, 50–64, 65–74)</td>
<td>Linear regression instead of Poisson regression was fitted because the denominators for the rates were not available</td>
<td>Not reported</td>
</tr>
<tr>
<td>Jørgensen and Gotzsche (2009)</td>
<td>Linear regression on prescreening incidence (1971–85)</td>
<td>Calendar year, stratified for three age class (30–49, 50–69, 70–84)</td>
<td>Linear regression instead of Poisson regression was fitted because the denominators for the rates were not available</td>
<td>Not reported</td>
</tr>
<tr>
<td>Jørgensen and Gotzsche (2009)</td>
<td>Linear regression on prescreening incidence (1980–94)</td>
<td>Calendar year, stratified for three age class (40–49, 50–69, 70–79)</td>
<td>Linear regression instead of Poisson regression was fitted because the denominators for the rates were not available</td>
<td>Not reported</td>
</tr>
<tr>
<td>Puliti et al. (2009)</td>
<td>Poisson regression on prescreening incidence (1986–90)</td>
<td>Age (annual, 40–79) and calendar year</td>
<td>Two-step Poisson analysis: 1) A first model (age, year) was fitted to the available prescreening incidence for each area 2) A pooled model (age, year, area) was fitted to prescreening incidence observed or estimated by model 1 (wherever missing)</td>
<td>1.012 (1.008–1.016)</td>
</tr>
<tr>
<td>Jørgensen et al. (2009)</td>
<td>Poisson regression on prescreening incidence (1971–1990)</td>
<td>Age class (5-years) and geographical area, stratified for screening age (50–69) and exceeded age (70–79)</td>
<td>The adjustment for prescreening geographical differences is not considered sufficient</td>
<td>1.0037 (1.0030–1.0045)</td>
</tr>
<tr>
<td>Duffy et al. (2010)</td>
<td>Poisson regression on prescreening incidence (1974–88)</td>
<td>Calendar year, stratified for age class (&lt;45, 45–49, 50–64, 65–69, 70+)</td>
<td>Plus adjustment for non-linear trends (dividing the expected numbers by the relative excess for &lt;45 years)</td>
<td>Not reported</td>
</tr>
</tbody>
</table>
Table 2 (Continued.)

<table>
<thead>
<tr>
<th>Paper</th>
<th>Model</th>
<th>Adjusted for . . .</th>
<th>Comments</th>
<th>Annual increase in the prescreening period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Martinez-Alonso et al. (2010)</td>
<td>Age-cohort model over all the study period</td>
<td>Age (annual, 25–84), year of birth, fertility rate and use of mammography</td>
<td>This model was used to estimate the background incidence in the absence of screening assuming the use of mammography equal zero</td>
<td>Not estimated</td>
</tr>
<tr>
<td>de Gelder et al. (2011)</td>
<td>Microsimulation modelling based on prescreening incidence data</td>
<td>Not necessary</td>
<td>Via micro simulation modelling the incidence without screening in a population aged 0–100 was predicted</td>
<td>Not estimated</td>
</tr>
</tbody>
</table>

Table 3a Details of the adjustment for lead time: papers with statistical adjustment

<table>
<thead>
<tr>
<th>Paper</th>
<th>Method used for adjustment for lead time</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paci et al. (2004)</td>
<td>Individually screen-detected cases were shifted forward (for age and calendar year) according to the distribution of lead time</td>
<td>Assuming an exponential distribution for the sojourn time with a mean of 3.7 and 4.2 years for 50–59 and 60–69 years old respectively</td>
</tr>
<tr>
<td>Jonsson et al. 2005</td>
<td>A period equal to the 65% of lead time was added to age at diagnosis for all cases diagnosed in the screening period</td>
<td>(a) Assuming a fixed lead time of 2.4, 3.7, 4.2 and 4.6 years for 40–49, 50–59, 60–69 and 70–74 respectively. (b) 65% is the proportion of screen-detected cases observed at age 40–74 years</td>
</tr>
<tr>
<td>Olsen et al. (2006)</td>
<td>A multistate model was fitted to incidence of a cohort of women screened at least once</td>
<td>Assuming an exponential distribution of incidence for the non-progressive preclinical screen-detectable cancers and a test sensitivity of 100%</td>
</tr>
<tr>
<td>Paci et al. (2006)</td>
<td>Individually screen-detected cases were shifted forward (for age and calendar year) according to the distribution of lead time</td>
<td>Assuming an exponential distribution for the sojourn time with a mean of 3.7 years for 50–59 years old and 4.2 years for 60–69 years old</td>
</tr>
<tr>
<td>Martinez-Alonso et al. (2010)</td>
<td>A probabilistic model (including background incidence, competitive risks, distribution of sojourn time, sensitivity and dissemination of screening) was fitted</td>
<td>Using the probabilistic model, the expected incidence due to lead time (i.e. incidence with screening assuming no overdiagnosis) was estimated</td>
</tr>
</tbody>
</table>

Type of adjustment for lead time (no adjustment, statistical adjustment or compensatory drop).

Table 2 provides details of the estimation of underlying breast cancer risk in the selected papers. For the 16 estimates, three, nine used prescreening age distribution without considering temporal trend, seven, eleven used extrapolation of prescreening trends, two, eight were geographically controlled, two, four used risk factor adjustment and two, twelve estimated incidence by internal modelling.

Table 3a gives details of the estimates using statistical adjustment for lead time and Table 3b gives details for those taking the compensatory drop approach.

Table 4 describes the measure of overdiagnosis used in each selected paper, including the definition of the numerator and the denominator (with the age range to which the denominator pertains).

Single papers included in this review

Contemporaneous comparison group and no adjustment for lead time

Peeters et al. eight calculated overdiagnosis 12 years after the start of a pilot screening programme in Nijmegen, the Netherlands. The control population was represented by women aged greater than 35 years, who were resident in a neighbouring city where no mass screening was performed during the same time period. Incidence rates in screened and unscreened areas in the period previous to the screening programme were observed to verify the comparability of the two populations. No adjustment for lead time was used.

Postponement of screen-detected cases

The two papers by Paci et al. nine, thirteen (Italy) assumed an exponential distribution of breast cancer sojourn time. The probability that each screen-detected case identified in the screening programme would have surfaced clinically in the subsequent years after detection was calculated. The sum of these probabilities over all screen-detected cases, year by year, gives an estimate of the number of screen-detected cases that would have arisen clinically in each calendar year. In the first paper, nine based on an evaluation of the service screening programme in Florence, the expected number of cases was estimated by applying the age-specific incidence rates observed before the start of the screening programme.
programme to the age distribution of the target population during the study period, without considering any temporal trend. In the later paper, the method of postponement of screen-detected cases was applied to a larger data-set which included various areas of central and northern Italy and the prescreening temporal trend was taken into account.

Jonsson et al. estimated overdiagnosis in the Swedish screening programme as the relative risk adjusted for lead time in the so-called ‘stabilized phase’ (from year 7 onwards). A period equal to 65% of the estimated age-specific lead time was added to age at diagnosis for all cases diagnosed in the screening period (65% was the proportion of screen-detected cases in the relevant age range and period).

The statistical adjustment for lead time used in the two papers by Paci et al. was different in several respects from that employed by Jonsson et al. Specifically, in the papers by Paci et al.:

1. The adjustment for lead time was applied to the individual screen-detected case;
2. An exponential distribution of the lead time was used to calculate the probability to surface clinically in each year;
3. Screen-detected cases were moved forward on both the age and calendar year axes.

Conversely, Jonsson et al. attributed the proportion of lead time to all cases according to the proportion of screen-detected cases in the population, added a fixed duration of lead time and moved cases forward on the age axis only. It should be noted that the occurrence of screen-detected cases replaces the ‘future incidence’ of cancers which would have occurred not only at a woman’s older age but also postponed in terms of calendar year. The translation of the diagnosis date along the calendar year axis for those cases which had not surfaced before the end of the study period moves them to a later time and, correctly, they are not included in the numerator of the overdiagnosis estimate.

Beyond the methodological differences, we suggest that the paper by Jonsson is a good example of what can happen in observational studies using an historical comparison which cannot be fully controlled. In a subsequent paper co-authored by Jonsson, it was stated that the results of the Swedish study reported in Jonsson et al. were not explicitly attributed to overdiagnosis. Other potential factors were given (including changes in risk factor adjustment for lead time).

Table 3b Details of the adjustment for lead time: papers using the compensatory drop method

<table>
<thead>
<tr>
<th>Paper</th>
<th>Age and period considered</th>
<th>Incidence excess in the screened group</th>
<th>Exceeded age for screening</th>
<th>% women who have had the opportunity to screen</th>
<th>Mean (range) follow-up after screening stops</th>
<th>Compensatory drop in the older age groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zahl et al. (2004)</td>
<td>50–69 years in the 1998–99</td>
<td>+54%</td>
<td>70–74 years in the 2000</td>
<td>80%</td>
<td>2.5 years (1–4 years)</td>
<td>−11%</td>
</tr>
<tr>
<td>Zahl et al. (2004)</td>
<td>50–69 years in the 1997–2000</td>
<td>+45%</td>
<td>(1) 70–74 years in 1997–2000</td>
<td>(1) 100% (continuing screening)</td>
<td>(1) No follow-up (2) 4.2 years (1–10 years)</td>
<td>(1) Not observed</td>
</tr>
<tr>
<td>Waller et al. (2007)</td>
<td>50–64 years in 1990–2001</td>
<td>+73% (first screen) + 18–35% (subsequent)</td>
<td>65–67 years in 1997–2000</td>
<td>100%</td>
<td>1–3 years</td>
<td>−12%</td>
</tr>
<tr>
<td>Jørgensen and Gottzsche (2009)</td>
<td>50–64 years in 1993–1999</td>
<td>+41%</td>
<td>68–70 years in 1993–1999</td>
<td>82%</td>
<td>4–6 years</td>
<td>−8%</td>
</tr>
<tr>
<td>Jørgensen and Gottzsche (2009)</td>
<td>50–69 years in 1998–2006</td>
<td>+35%</td>
<td>70–84 years in 1998–2006</td>
<td>94% (1/3 continuing screening)</td>
<td>2.3 years (0–15 years)</td>
<td>−10%</td>
</tr>
<tr>
<td>Jørgensen and Gottzsche (2009)</td>
<td>50–69 years in 2000–2006</td>
<td>+42%</td>
<td>70–79 years in 2000–2006</td>
<td>75%</td>
<td>3.9 years (1–10 years)</td>
<td>−15%</td>
</tr>
<tr>
<td>Puliti et al. (2009)</td>
<td>60–69 years in 1990–1999</td>
<td>+21%</td>
<td>70–83 years in 1991–2004 (belong to the cohort)</td>
<td>100%</td>
<td>4.7 years (1–14 years)</td>
<td>−13%</td>
</tr>
<tr>
<td>Jørgensen et al. (2009)</td>
<td>50–69 years in 1991–2003</td>
<td>+40%</td>
<td>70–79 years in 1998–2003</td>
<td>92%</td>
<td>4.6 years (1–10 years)</td>
<td>−10%</td>
</tr>
<tr>
<td>Duffy et al. (2010)</td>
<td>45–64 years in 1989–2003</td>
<td>+13%</td>
<td>65–74 years in 1989–2003</td>
<td>60%</td>
<td>5.0 years (1–15 years)</td>
<td>−8.1%</td>
</tr>
<tr>
<td>De Gelder et al. (2011)</td>
<td>0–74 years in 2006</td>
<td>+7%</td>
<td>75–100 years in 2006</td>
<td>79%</td>
<td>6.1 years (1–16 years)</td>
<td>−11.7%</td>
</tr>
</tbody>
</table>

\(^1\)These figures have been estimated on the basis of data reported in the papers (start year of screening and target age of screening). For reference 10b (Sweden), we assumed that age of screening was extended to 70–74 years in the 1995.\(^2\)

\(^2\)It should be noted that these compensatory drops (11% in Norway and 12% in Sweden) were not considered to estimate overdiagnosis.
Table 4  Details of the measure of overdiagnosis

<table>
<thead>
<tr>
<th>Paper</th>
<th>Measure of overdiagnosis (OD)</th>
<th>Numerator</th>
<th>Denominator</th>
<th>Age range (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peeters et al. (1989)†</td>
<td>Excess incidence/Expected incidence</td>
<td>The difference between the observed incidence in screened area and the observed incidence in unscreened area adjusted for birth year</td>
<td>The observed incidence in the unscreened area adjusted for birth year</td>
<td>35+</td>
</tr>
<tr>
<td>Paci et al. (2004)</td>
<td>Adjusted excess incidence/Expected incidence</td>
<td>The difference between the post-screening incidence statistically adjusted for lead time and the prescreening incidence adjusted for age</td>
<td>Prescreening incidence adjusted for age</td>
<td>50–84</td>
</tr>
<tr>
<td>Zahl et al. (2004)</td>
<td>Excess incidence/Expected incidence</td>
<td>The difference between the post-screening incidence and the prescreening incidence adjusted for age</td>
<td>Prescreening incidence adjusted for age</td>
<td>50–69</td>
</tr>
<tr>
<td>Zahl et al. (2004)</td>
<td>Excess incidence/Expected incidence</td>
<td>The difference between the post-screening incidence and the prescreening incidence adjusted for age</td>
<td>Prescreening incidence adjusted for age</td>
<td>50–69</td>
</tr>
<tr>
<td>Jonsson et al. (2005)</td>
<td>Adjusted excess incidence/Expected incidence</td>
<td>The difference between the post-screening incidence statistically adjusted for lead time and the prescreening incidence adjusted for age, temporal trend and area</td>
<td>Prescreening incidence adjusted for age, temporal trend and area</td>
<td>40–74</td>
</tr>
<tr>
<td>Olsen et al. (2006)</td>
<td>Adjusted excess incidence/Observed incidence</td>
<td>The adjusted excess incidence was estimated by multistate modelling</td>
<td>Observed incidence among screened</td>
<td>50–69 followed for 6 years</td>
</tr>
<tr>
<td>Paci et al. (2006)</td>
<td>Adjusted excess incidence/Expected incidence</td>
<td>The difference between the post-screening incidence statistically adjusted for lead time and the prescreening incidence adjusted for age, temporal trend and area</td>
<td>Prescreening incidence adjusted for age, temporal trend and area</td>
<td>50–74</td>
</tr>
<tr>
<td>Waller et al. (2007)</td>
<td>Absolute increase of lifetime risk of breast cancer due to screening</td>
<td>The difference between the lifetime risk of breast cancer with screening and the lifetime risk of breast cancer without screening (as estimated by APC model)</td>
<td>Not applicable (The measure was expressed as absolute increase)</td>
<td>–</td>
</tr>
<tr>
<td>Jørgensen and Gotzsche (2009)</td>
<td>Adjusted excess incidence/Expected incidence</td>
<td>The difference between the post-screening incidence (no compensatory drop was found) and the prescreening incidence adjusted for age and temporal trend</td>
<td>Prescreening incidence adjusted for age and temporal trend</td>
<td>50–64</td>
</tr>
<tr>
<td>Jørgensen and Gotzsche (2009)</td>
<td>Adjusted excess incidence/Expected incidence</td>
<td>The difference between the post-screening incidence corrected for compensatory drop and the prescreening incidence adjusted for age and temporal trend</td>
<td>Prescreening incidence adjusted for age and temporal trend</td>
<td>50–69</td>
</tr>
<tr>
<td>Jørgensen and Gotzsche (2009)</td>
<td>Adjusted excess incidence/Expected incidence</td>
<td>The difference between the post-screening incidence corrected for compensatory drop and the prescreening incidence adjusted for age and temporal trend</td>
<td>Prescreening incidence adjusted for age and temporal trend</td>
<td>50–69</td>
</tr>
<tr>
<td>Puliti et al. (2009)</td>
<td>Adjusted excess incidence/Expected incidence</td>
<td>The difference between the post-screening incidence corrected for compensatory drop and the prescreening incidence adjusted for age and temporal trend</td>
<td>Prescreening incidence adjusted for age and temporal trend</td>
<td>60–69 followed for 15 years</td>
</tr>
<tr>
<td>Jørgensen et al. (2009)</td>
<td>Adjusted excess incidence/Expected incidence</td>
<td>The difference between the post-screening incidence corrected for compensatory drop and the prescreening incidence adjusted for age and temporal trend</td>
<td>Prescreening incidence adjusted for age and temporal trend</td>
<td>50–69</td>
</tr>
<tr>
<td>Duffy et al. (2010)</td>
<td>No. cases overdiagnosed for 1000 women screened for 20 years</td>
<td>The difference between the post-screening incidence corrected for compensatory drop and the prescreening incidence adjusted for age and temporal trend</td>
<td>Person-years of screened women</td>
<td>–</td>
</tr>
<tr>
<td>Martínez-Alonso et al. (2010)</td>
<td>Adjusted excess incidence/Expected incidence</td>
<td>The difference between the observed incidence with screening and the expected incidence with screening due to lead time (assuming no overdiagnosis)</td>
<td>The expected incidence with screening due to lead time</td>
<td>40+</td>
</tr>
<tr>
<td>de Gelder et al. (2011)</td>
<td>Adjusted excess incidence/Expected incidence</td>
<td>The difference between the predicted incidence with screening and predicted incidence without screening (as estimated by MISCAN)</td>
<td>Predicted incidence without screening (as estimated by MISCAN)</td>
<td>0–100</td>
</tr>
</tbody>
</table>
prevalence such as hormone replacement therapy). For this reason, the estimates of this paper were considered not fully adjusted for breast cancer risk.

Other types of statistical adjustment for lead time

Olsen et al.\(^2\) (Denmark) estimated the natural history of breast cancer by multistate modelling in a similar approach to that of Day and Walter.\(^2\) The model included the incidence of truly progressive preclinical cancers, the time spent in the preclinical state, the screening test sensitivity and the incidence of non-progressive preclinical (and therefore overdiagnosed) cancers. The authors estimated these parameters from the data on screen-detected and interval cancers. Sensitivity analyses were carried out, varying the screening sensitivity. The authors concluded that 4.8% of all cancers diagnosed among participants during the first two rounds were overdiagnosed. To make this estimate comparable with the others, we re-calculated it as the percentage of the expected incidence in the absence of screening, as the following: the absolute number of overdiagnosed cases was estimated as 30 (0.048 × 627) and the number of expected cases in the absence of screening was estimated by applying the underlying breast cancer incidence to the observed person years during the first two rounds (0.0038 × 112,860 = 429), obtaining an estimate of 7.0% (30/429).

Martinez-Alonso et al.\(^1\) used a probabilistic model taking into account background incidence, competing risks, the distribution of sojourn time in the preclinical state, mammographic sensitivity and the dissemination of screening in Catalonia (Spain) to estimate the increased age-specific incidence due to lead time. Overdiagnosis was estimated as the difference between the observed incidence with screening and the modelled incidence taking lead time into account. The authors modelled the background incidence of breast cancer during the period 1980–2004 using an age-cohort model where the cohort effects were split into three components: fertility rate, percentage of women undergoing mammography at age 50 and year of birth. Breast cancer incidence in the absence of screening was derived from this model by considering that the proportion of women having mammograms at age 50 was zero. It should be noted that all temporal effects were attributed to ageing and cohort characteristics, without an additional period effect. This analytical approach resulted in a wide variability of the estimates by birth cohort. While no overdiagnosis was attributed to the oldest cohort born in 1935, the estimate of overdiagnosis was almost 50% in the youngest cohort born in 1950. This variability was not adequately explained by the authors. In addition, the selected birth cohorts differ significantly from each other in relation to screening exposure. Women born in 1935 had been screened from age 55 to 64 and followed up until 69 years, whereas women born in 1950 were followed up only to age 54, so that their exposure represented mostly prevalence screens, and they had no postscreening observation.

Studies that took into consideration the compensatory drop

Zahl et al.\(^1\) followed a dynamic population approach, and found a breast cancer reduction in the older age groups both in Sweden and Norway (12% and 11% respectively) but these findings were not incorporated into the estimate of overdiagnosis because the results were not statistically significant. Indeed, the estimate of overdiagnosis is the same as the incidence excess observed in the screening age group (see Table 3b). For this reason, the paper was included in Table 3b together with all other papers which took into consideration the compensatory drop method, but was presented as having ‘no adjustment for lead time’ in Table 1. In addition, the authors did not clearly explain how they adjusted for breast cancer risk. The estimate of the annual percent change was reported in the tables of the Results section but it was not taken into account in estimating overdiagnosis (see also Table 2).

In the paper by Waller et al.\(^1\) (England and Wales), a dynamic population was analysed using a model including age, period and cohort parameters, indicator variables for screening (initial screen, successive screens and different periods after screening) and use of hormone replacement therapy. This analysis, previously proposed by Moller et al.,\(^2\) allowed the authors to interpret in a longitudinal way the dynamic population data, ensuring that the deficit in incidence was measured for women who had had the opportunity to be screened. There may be bias in the estimates, however, arising from modelling aggregate proportions and interpreting results from these as effects at individual level.\(^\) Waller and colleagues measured overdiagnosis as the absolute increase of lifetime risk of breast cancer due to screening. In order to make it comparable with the other reviewed papers (where it is expressed as a percentage of the expected incidence in the absence of screening), we recalculated the estimate dividing the lifetime risk of breast cancer with screening by the lifetime risk of breast cancer without screening (8.6%/7.8% = 1.10).

Jørgensen and Gotzsche\(^1\) performed a linear regression of time on incidence in the prescreening and screening periods separately (the latter after a prevalence peak), from screening programmes in several countries. In order to estimate both the excess in the screened age range and the drop in ages above the screening age range, the rate ratios between the result for the last observation year determined by linear regression and the expected incidence in that year were calculated.

The analysis was not performed on actual data obtained from cancer registries, but on data extracted from selected papers and, in at least one paper, from a graphic illustration (both authors extracted data independently, with differences resolved by discussion). The authors used simple linear regression rather than Poisson regression, to estimate breast cancer trends, because the denominators for the rates were not available, and, therefore, the confidence interval of the estimates could not be calculated. The authors modelled the observed incidence in the postscreening period. They used only the last year determined by linear regression rather than using the available observed cumulative incidence in the postscreening period. This introduced further statistical uncertainty relating to the specific trend modelled in the postscreening period. The use of the modelled rates referring to one year only, instead of the observed cumulative incidence, does not take into account the temporal duration of both the excess and the drop. The compensatory drop was therefore not correctly
estimated, and indeed no drop was applied in the UK case. In addition, if the levels of breast cancer incidence rates increased abruptly in the years immediately before the introduction of screening, the authors excluded those years from estimates of trends before screening. In the case of the UK, they excluded years 1985–1988. The choice of the reference period for prescreening incidence had an impact on the resulting prescreening trend and, therefore, on the expected incidence in both the screening and in the exceeded age range (see Table 3b). The overall effect of the Jørgensen and Gotzsche approach in removing years 1985–1988 (the years of highest incidence in their prescreening period) and calculating overdiagnosis only for year 1999 (excluding the years of lowest incidence in their screening period) was to increase the estimated overdiagnosis.

Pulli et al.16 (Italy) used a cohort approach. The authors followed a birth cohort (women aged 50–69 years at the beginning of service screening) for 15 years and used the breast cancer reduction observed in the period after the last screening to adjust for lead time. The follow-up period was long enough to take into account the lead time and to provide a correct estimate of overdiagnosis only for women aged 60–69 at entry. The expected incidence was estimated by modelling the pre-screening incidence by age and calendar year. A sensitivity analysis assuming no trend was also performed.

In another study by Jørgensen et al.17 breast cancer incidence in two Danish areas (Copenhagen and Funen), in which screening took place, was compared with incidence in the rest of Denmark where there was no screening in the same time period. The authors stated that, because they compared screened and non-screened regions, general changes in the background incidence would not materially affect the estimate of overdiagnosis. Nevertheless, in such a comparative study, the adjustment for geographical differences is the crucial term for the validity of the study. Authors reported that prescreening incidence rates (1971–1990) in the screened area were higher than in the not screened area (average 214 versus 198 breast cancers per 100,000 person years for women aged 50–69). On the contrary, the ratio of screened versus unscreened area (pre-screening) used in the Poisson regression is 0.90, showing an inverse relation. In Table 1 of this paper,17 the incidence rates in the screened areas were 214 per 100,000 in the period 1971–1990 and 392 in 2001–2003. In the non-screened areas, rates of 198 and 314 per 100,000 were observed, giving a relative risk of

\[ RR = \frac{392 \times 198}{214 \times 314} = 1.16 \]

i.e., a 16% excess. The corresponding figures for ages 70–79 give a relative risk of

\[ RR = \frac{327 \times 264}{273 \times 367} = 0.86 \]

Thus there is a compensatory drop in the upper age group of similar relative size (although smaller in absolute terms) to the excess in the 50–69 year age group. It should also be noted that between the prescreening period, 1971–1990, and the screening period, 1991–2003, incidence increased in the 35–49 year age group, suggesting that some of the observed incidence in the screened age group is independent of the screening.

In a previous publication,25 regional differences in breast cancer incidence in Denmark were assessed over a 20-year prescreening period (1970–1989). The study showed important regional differences with an incidence in the municipality of Copenhagen significantly higher than in the rest of Denmark. Therefore, due to all the reservations above, the adjustment for breast cancer risk cannot be considered sufficient.

Duffy et al.18 (England) estimated the temporal trend in incidence from 1974 to 1988 before the start of service screening and projected this to estimate the expected incidence in 1989–2003. The authors also adjusted for any non-linear trends comparing the expected and observed incidence relative to women aged <45, in which very little screening took place. Overdiagnosed cases were calculated as the number of excess cases in the 45–49 and 50–64 years old age groups minus the deficit in the 65–69 and 70+ years old age groups. The estimate of overdiagnosis was reported as the number of overdiagnosed cases for 1000 women screened for 20 years. From the data reported in the paper, we recalculated it as the net excess of breast cancer cases divided by the number of expected cases in the age range 45–64 in the absence of screening (6061/186,173 = 0.033).

In de Gelder et al.20 (the Netherlands) the observed breast cancer incidence between 1990 and 2006 was taken into account by the MISCAN microsimulation model and the natural history of breast cancer was modelled assuming specific transitional probabilities between different states. Observed breast cancer incidence in the presence of screening was modelled and compared with the predicted incidence without screening. The overdiagnosed cases were estimated by comparing the number of excess breast cancers in women of screening age with the number of deficit breast cancers in the group exceeding the screening limit in a steady-state screening situation.

Summary of the estimates of overdiagnosis

Because methodological approaches used to estimate overdiagnosis differ between studies, and there is little agreement in the way the data should be analysed, a formal meta-analysis of the estimates would be inappropriate. We classified the estimates according to the adjustment for breast cancer risk and lead time bias, as these are fundamental to an accurate assessment of overdiagnosis. We classified the following studies as having estimates of overdiagnosis that were not adequately adjusted for breast cancer risk: Paci et al.,9 Zahl et al.,10 Jonsson et al.,11 Jørgensen et al.17 and Martinez-Alonso et al.19 Secondly, the estimates from the papers by Peeters et al.,8 Zahl et al.10 and by Jørgensen and Gotzsche15 were classified as not adequately adjusted for lead time. On this basis, the estimates of overdiagnosis adjusted for breast cancer risk and lead time bias 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 (from Table 1 or estimated in the sections above). No reliable estimates were available for Norway, Sweden or Spain. The unadjusted or incompletely adjusted estimates ranged from 0% to 54%. Fig. 1 shows the estimates of overdiagnosis.
DISCUSSION

The methodological framework used in this review for the evaluation of overdiagnosis estimates in observational studies is based on identifying the two main potential biases that can affect the estimates. Overdiagnosis can correctly be estimated by comparing incidence in screened and unscreened populations, provided that the underlying risks of breast cancer in these two groups are similar, and that the effect of lead time is accounted for.26

In adjusting for breast cancer risk, using the correct estimate of temporal trend is crucial when data are derived from non-concurrent screened and unscreened populations. The importance of this aspect can be appreciated by comparing the estimates from very similar population data presented by Duffy et al.16 and by Jørgensen and Gotzsche.15 Both groups analysed incidence before and after screening was introduced in the UK. As noted above, Jørgensen and Gotzsche probably underestimated the expected incidence in the screening period because they excluded the four years before the implementation of screening when estimating the prescreening trend. This could explain why they found no compensatory drop, whereas Duffy et al. did. This issue has been extensively dealt with by Kopans et al.27 in a recent publication.

The so-called ‘compensatory drop’ method is commonly used for adjusting for lead time.27 In the absence of overdiagnosis, the initial increase in breast cancer incidence in the screened group would be fully compensated for by a similar decrease in cancers among older age groups no longer offered screening. The compensatory drop method can be applied both in the analysis of the dynamic population and in cohort studies evaluating a group of people defined by year of birth or by individual enrolment. For cohort studies, a valid estimate of overdiagnosis can be obtained by comparing the cumulative incidence between screened and unscreened women after a sufficient follow-up time.5 In the case of a dynamic population, the excess incidence is calculated using the screening age group during the screening period and the compensatory drop is estimated among women whose age is above that for screening. Therefore it is crucial to check if, and how many, women in the older age group have really had the opportunity to be screened. This condition is implied by definition in a cohort approach. The majority of observational studies estimated breast cancer overdiagnosis using temporal trends or geographical differences in breast cancer incidence in a dynamic population. Among all the selected papers, only those by Olsen et al.12 and by Puliti et al.16 used the cohort approach.

Figure 1: Overdiagnosis estimates classified according to the presence/absence of both the adjustments. The numbers indicate the related reference. Notes: (1) For the paper by Jonsson et al.,11 we reported the pooled estimate for 40–74 years (20%) calculated by Jonsson himself. (2) For the paper by Martinez-Alonso et al.,19 we reported the estimate of the cohort of women born in 1950 considered by the authors themselves to be the best estimate (personal communication).
The compensatory drop method, both in a cohort and in a dynamic population, needs sufficient follow-up after screening stops to achieve a full adjustment for lead time. It has been shown\textsuperscript{20} that the estimate of overdiagnosis may further decrease as the number of women contributing to the deficit in incidence continues to increase. A compensatory drop in incidence is fully observed only if all women in the age group above the screening age have been invited to screening when they were in the eligible age range.

Another important consideration is the measure of overdiagnosis. As shown by de Gelder et al.,\textsuperscript{20} the estimate of overdiagnosis is strongly dependent on the denominator used to define the population at risk. If overdiagnosis is calculated as a relative risk for women of screening age, it could be almost double the estimate if women of all ages are included. There is no consensus about the measure to use, but the choice of the denominator should depend on the purpose of the overdiagnosis estimate. When comparing two or more estimates of overdiagnosis, or when the estimate of overdiagnosis is used in a balance of harms and benefits, it is crucial to specify to which population the estimates apply.

The variability in overdiagnosis estimates can also partly be explained by differences in screening policies and different uptake between programmes. All estimates considered in this review, except those by Olsen et al.\textsuperscript{12} and by Waller et al.,\textsuperscript{14} pertain to the screening target population, not to women actually screened, and they therefore strongly depend on screening compliance. Further, the extent of overdiagnosis may be affected by the intensity of screening (including screening interval and recall practice) and by the screening age range, both because of variation in the natural history of the disease with age, and because of increased mortality from other causes in older women.

Lastly, overdiagnosis estimates depend on the length of the screening period considered. It should be noted that some estimates\textsuperscript{12,13,16} pertain to the first two or three screening rounds only, including the prevalence screen. There is consistent evidence that the overdiagnosis rate is higher at the prevalence screen than in subsequent rounds.\textsuperscript{6,12} Therefore it is expected that these estimates would have been lower if they had pertained to the whole screening period (10 rounds over 20 years).

**CONCLUSION**

Estimation of the underlying expected incidence in the absence of screening is crucial to obtaining reliable estimates of overdiagnosis. When considering the adjustment for changes in breast cancer risk, we highlight the importance of the estimate of the annual percentage increase in the pre-screening period to determine the expected incidence when there is no contemporaneous control. We advocate that the annual increase should be explicitly reported in future papers, with sensitivity analyses reporting different estimates of overdiagnosis under different assumptions in the modelling of the expected incidence trend.

In adjusting for lead time, the compensatory drop method focuses on an actual observed incidence reduction, whereas the statistical adjustment method strongly depends on the assumptions of the model used (in particular the lead time distribution). However, the distribution of lead time can be estimated rigorously if detailed observations on incidence, stages, screen-detected cases and interval cancers are available.\textsuperscript{28} In principle, the cohort approach is preferable to the analysis of a dynamic population, because it follows the experience of a group of women who have truly had the opportunity to be screened, and allows an accurate evaluation of whether there is a sufficient follow-up after the last screen. If a cohort approach is not possible, an age-period-cohort analysis, which also includes indicator variables for the different phases of screening (prevalence screen, successive screens, and period after screening), as in the paper by Moller et al.,\textsuperscript{23} is recommended.

Analysis of the selected papers in this review and of the potential biases that may affect the estimates suggests that the most plausible estimates of overdiagnosis, expressed as a percentage of the expected incidence in the absence of screening, are relatively low, ranging from 1% to 10%, and that substantially higher estimates reported in the literature are likely to be overestimates of overdiagnosis due to lack of adjustment for breast cancer risk and/or lead time.

**EUROSCREEN WORKING GROUP**

**Coordinators:**

E Paci, M Broeders, S Hofvind, SW Duffy

**Members:**

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

**Affiliations:**

\textsuperscript{1}Ministère du travail de l’emploi et de la santé, Paris, France
\textsuperscript{2}CPO-Piedmont, Turin, Italy
\textsuperscript{3}Navarra Breast Cancer Screening Programme, Pamplona, Spain
\textsuperscript{4}S.C. Epidemiologia, ASL di Milano, Italy
\textsuperscript{5}Radboud University Nijmegen Medical Centre & National Expert and Training Centre for Breast Cancer Screening, Nijmegen, The Netherlands
\textsuperscript{6}Veneto Tumor Registry, Padua, Italy
\textsuperscript{7}Erasmus MC, Dept. Of Public Health Rotterdam, The Netherlands
\textsuperscript{8}Wollson Institute of Preventive Medicine, Queen Mary University of London, UK
\textsuperscript{9}Cancer Registry of Norway, Research Department and Oslo and Akershus University College of Applied Science, Oslo, Norway
\textsuperscript{10}Department of Radiation Sciences, Oncology, Umeå University, Umeå, Sweden
\textsuperscript{11}Centre for Epidemiology and Screening, University of Copenhagen, Denmark
ACKNOWLEDGEMENTS

The 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


APPENDIX

Search Strategy

We searched the National Library of Medicine Pubmed up to February, 2011 using the following search strategies:

1. overdiagnosis
2. mammography screening
This search strategy retrieved a total of 99 papers.

'Mass Screening'[Mesh]
'Mammography'[Mesh]
'Breast Neoplasms'[Mesh]
'Diagnostic Errors'[Mesh]
'False Positive Reactions'[Mesh]
'Reproducibility of Results'[Mesh]
'Sensitivity and Specificity'[Mesh]

(((#7) OR #8) OR #9) OR #10

(((#4) AND #5) AND #6) AND #2

(#12) AND #11

(#7) OR #8

(#12) AND #14

(#14) OR #1

(#12) AND #16

((#13) AND #15) AND #17

(#18) NOT #3

This search strategy retrieved a total of 382 papers.

((#13) OR #15) OR #17

(#20) NOT #3

This search strategy retrieved a total of 1040 papers.

((#4 AND #6) AND #2

(((#1) OR #7) OR #8) OR #9) OR #10

(#22) AND #23

This search strategy retrieved a total of 1168 papers.

((#1) AND #4) AND #6

This search strategy retrieved a total of 83 papers.

harm and benefit

(#26) AND #2

This search strategy retrieved a total of 37 papers.

advantages and disadvantages

(#28) AND #2

This search strategy retrieved a total of 45 papers.

'Incidence'[Mesh]

((#30) AND #4) AND #6

This search strategy retrieved a total of 596 papers.

Publications of authors expert in the field:

Duffy S[Author] AND #1

This search strategy retrieved a total of 15 papers.

Paci E[Author] AND #1

This search strategy retrieved a total of 6 papers.

Lynge E[Author] AND #1

This search strategy retrieved 8 articles.

Zahl PH[Author] AND #1

This search strategy retrieved 8 articles.

PubMed ‘related articles’ to the following references suggested by experts in the field:


This function retrieved a total of 240 papers.


This function retrieved a total of 104 papers.


This function retrieved a total of 95 papers.


This function retrieved a total of 106 papers.


This function retrieved a total of 119 papers.


This function retrieved a total of 113 papers.

Jørgensen KJ. Mammography screening is not as good as we hoped. Maturitas January 2010;65(1):1–2

This function retrieved a total of 138 papers.


This function retrieved a total of 115 papers.


This function retrieved a total of 174 papers.


This function retrieved a total of 124 papers.


This function retrieved a total of 150 papers.

References from the following published articles:

- Welch HG. Overdiagnosis and mammography screening. BMJ 9 July 2009;339
• Ciatto S. The overdiagnosis nightmare: a time for caution. *BMC Womens Health* 16 December 2009;9:34

In addition we consulted the following publication:
Osservatorio Nazionale Screening Ottavo Rapporto 2009
http://www.osservatorionazionalescreening.it/ita/images/stories/8_Rapporto_ONS.pdf

These searches were supplemented with suggestions by experts in the field.

We considered all articles published in English language up to February 2011 (no date restriction). We imported into ProCite all articles and we selected the papers considered relevant after the reading of title and abstracts.