

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 EUROSREEN Working Group (members listed at the end of the paper)

J Med Screen 2012;19 Suppl 1:42–56
DOI: 10.1258/jms.2012.012082

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.

See end of article for authors' affiliations

Correspondence to:
Eugenio Paci, Clinical and Descriptive Epidemiology Unit, ISPO – Cancer Research and Prevention Institute, Via delle Oblate 2, 50141 Florence, Italy; e.paci@ispo.toscana.it

Accepted for publication
21 June 2012

INTRODUCTION

Overdiagnosis is defined as the detection of a breast cancer at screening, histologically confirmed, that would never have been identified clinically in the lifetime of the woman. Some overdiagnosis is likely to result from mammographically detected cancers that may have remained asymptomatic throughout a woman's lifetime. Other cases are overdiagnosed because, although they are detected early, the woman dies of other causes before symptoms would have developed. The woman would only experience the harmful effects of early diagnosis and treatment without the opportunity to benefit. Due to the invasive procedures involved and the physical and psycho-social impact of the treatment, overdiagnosis may be considered the most adverse outcome associated with mammographic screening. Unfortunately, it is not possible to recognize which individual cases of breast cancer actually result from overdiagnosis. The number can only be estimated at population level based on analysis of data collected over years of screening. Several studies have tried to

quantify overdiagnosis of breast cancer but estimates vary widely and debate exists over the rate.^{1–4}

The paradigm for estimating overdiagnosis is to compare the cumulative incidence of breast cancers in the intervention and control arms several years after screening ends using data from a randomized controlled trial of screening in which the control group was not offered screening at the end of the trial.^{5–7} Moss⁶ estimated the overdiagnosis of breast cancer separately for randomized trials with or without screening in the control arm at the end of the trial. Among trials in which the control group was not offered screening, the two Canadian trials estimated 14% and 11% excess of breast cancers in the intervention arm eight years after the end of trial. In the evaluation of the extended follow-up of the Malmo trial, Zackrisson *et al.*⁷ estimated that overdiagnosis 15 years after the end of trial was 10% for all cases and 7% for invasive breast cancers.

However, the randomized trial estimates refer to an experience of mammographic screening in an experimental

setting 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 studies^{8–20}, 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 time^{5,13}). 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 1 Characteristics and main results of the studies of overdiagnosis

Paper	Population			Intervention		Comparison			Outcomes		
	Country	Calendar period	Type of population	Age and interval of screening	Start year of screening*	Reference population	Adjustment for breast cancer risk	Adjustment for lead time	Measure of overdiagnosis (OD)†	Estimate OD (only invasive)	Estimate OD (in situ and invasive)
Peeters <i>et al.</i> (1989) ⁸	The Netherlands (Nijmegen)	1970–1986	Dynamic population	35+ years biennial	1975	Incidence of neighbouring unscreened area	Birth year	No adjustment	Excess incidence/Expected incidence	Not reported	11%
Paci <i>et al.</i> (2004) ⁹	Italy (Florence)	1985–1999	Dynamic population	50–69 years biennial	1990	Prescreening incidence	Age	Statistical adjustment	Adjusted excess incidence/Expected incidence	1%	5%
Zahl <i>et al.</i> (2004) ¹⁰	Norway (AORH counties)	1971–2000	Dynamic population	50–69 years biennial	1996	Prescreening incidence	Age	No adjustment†	Excess incidence/Expected incidence	54%	Not reported
Zahl <i>et al.</i> (2004) ¹⁰	Sweden	1971–2000	Dynamic population	Different age range: the broadest 40–74 the less extensive 50–69	1986–1990	Prescreening incidence	Age	No adjustment†	Excess incidence/Expected incidence	45%	Not reported
Jonsson <i>et al.</i> (2005) ¹¹	Sweden (11 counties)	1971–2000	Dynamic population	Different age range: the broadest 40–74 the less extensive 50–69	1986–1990	Prescreening incidence	Age, temporal trend and area	Statistical adjustment	Adjusted excess incidence/Expected incidence	0–54% according to age	Not reported
Olsen <i>et al.</i> (2006) ¹²	Denmark (Copenhagen)	1991–1996	Cohort by enrolment	50–69 years biennial	1991	Incidence among screened women	Not necessary	Statistical adjustment	Adjusted excess incidence/ Observed incidence	Not reported	4.8%
Paci <i>et al.</i> (2006) ¹³	Italy (Northern and Central areas)	1986–2001	Dynamic population	50–69 years biennial	1991–1997	Prescreening incidence	Age, temporal trend and area	Statistical adjustment	Adjusted excess incidence/Expected incidence	3.2%	4.6%
Waller <i>et al.</i> (2007) ¹⁴	England and Wales	1971–2001	Dynamic population	50–64 years triennial (extended to 65–70)	1988 (2002)	Pre- and postscreening incidence	Age, period, birth cohort and use of HRT	Compensatory drop	Absolute increase of lifetime risk of breast cancer	0.86%	Not reported
Jørgensen and Gotzsche (2009) ¹⁵	England and Wales	1971–1999	Dynamic population	50–64 years triennial (extended to 65–70)	1988 (2002)	Prescreening incidence	Age and temporal trend	Compensatory drop	Adjusted excess incidence/Expected incidence	41%	57% (assuming 10% CIS)
Jørgensen and Gotzsche (2009) ¹⁵	Sweden	1971–2006	Dynamic population	Different age range: the broadest 40–74 the less extensive 50–69	1986–1990	Prescreening incidence	Age and temporal trend	Compensatory drop	Adjusted excess incidence/Expected incidence	31%	46% (assuming 10% CIS)
Jørgensen and Gotzsche (2009) ¹⁵	Norway (AORH counties)	1980–2006	Dynamic population	50–69 years biennial	1996	Prescreening incidence	Age and temporal trend	Compensatory drop	Adjusted excess incidence/Expected incidence	37%	52% (assuming 10% CIS)

Author (Year)	Country	Birth cohort	Age group	Year	Prescreening incidence	Age and temporal trend	Compensatory drop	Adjusted excess incidence/Expected incidence	Other
Puliti <i>et al.</i> (2009) ¹⁶	Italy (Florence)	1986–2004	50–69 years biennial	1990	Prescreening incidence	Age and temporal trend	Compensatory drop	Adjusted excess incidence/Expected incidence	0%
Jorgensen <i>et al.</i> (2009) ¹⁷	Denmark (Funen and Copenhagen)	1971–2003	50–69 years biennial	1991–93	Incidence of neighbouring un-screened area	Age and geographical area	Compensatory drop	Adjusted excess incidence/Expected incidence	Not reported
Duffy <i>et al.</i> (2010) ¹⁸	England	1974–2004	50–64 years triennial (extended to 65–70)	1988 (2002)	Prescreening incidence	Age and temporal trend	Compensatory drop	N° cases overdiagnosed for 1000 women screened for 20 years	2.3
Martinez-Alonso <i>et al.</i> (2010) ¹⁹	Spain (Catalonia)	1980–2004	50–64 years biennial (extended to 65–69)	1990 (2000)	Pre- and post-screening incidence	Age, year of birth, fertility rate and use of mammography	Statistical adjustment	Adjusted excess incidence/Expected incidence	0.4%–46.6% according to birth cohort
de Gelder <i>et al.</i> (2011) ²⁰	The Netherlands	1989–2006	50–69 years biennial (extended to 70–74)	1990 (1999)	Predicted incidence without screening by MISCAN	Not necessary	Compensatory drop	Adjusted excess incidence/Expected incidence	Not reported

*The year within brackets refers to the start of screening for the age class reported in brackets in the previous column. The range 1991–1997 for the paper by Paci *et al.*¹³ indicates the range of beginning of screening among area included in the study. The range 1991–1993 for the paper by Jorgensen *et al.*¹⁷ indicates that screening started in 1991 in Copenhagen and in 1993 in Funen

[†]Adjusted 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

Measure of overdiagnosis

Overdiagnosis has been reported using different epidemiological measures. The numerator is the absolute number of overdiagnosed cases estimated as the residual excess of breast cancer cases after considering adjustments for lead time and for breast cancer risk. This estimate of the absolute excess of breast cancer cases is usually compared with the cumulative number of cases expected in the same temporal period in the absence of screening in a certain age range. The estimated overdiagnosed cases can be expressed relative to a variety of denominators, including expected cases in the screening age range or lifetime, observed cases detected in the screened or invited population or screen-detected cancers. The choice of denominator will affect the size of the estimated rate and its interpretation.

We expressed estimates of overdiagnosis from each paper as a percentage of the expected incidence in the absence of screening, in order to make the estimates more comparable. However, the variability in the age range to which the denominator pertains could not be removed using the available data and the range is therefore reported in the tables.

'Screened' and 'unscreened' populations

Overdiagnosis is estimated by comparing incidence in screened and unscreened populations (after adjusting for lead time bias and breast cancer risk). However, the terms 'screened' and 'unscreened' can be confusing.

In almost all the papers we considered, the nominal 'screened' population was defined as the screening age classes and the calendar years after screening began. Therefore 'screened' actually means 'having the opportunity to be screened' because not all the women of the target population were actually invited to screening (for example, during the implementation phase) and only a proportion of invited women are actually screened (compliance). Only the papers by Olsen *et al.*¹² and by Waller *et al.*¹⁴ apply to women actually screened.

The incidence in the absence of screening is usually (but not invariably – see the papers by Peeters *et al.*⁸ and Jørgensen *et al.*¹⁷) not estimated directly from a contemporaneous 'unscreened' population, but indirectly, for example by extrapolation of incidence trends from a prescreening period. The incidence in the 'unscreened' population is estimated by different methods in the selected papers, as reported in Table 2.

RESULTS

We included 13 primary studies in our review, reporting 16 estimates 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 in order of year of publication. When the paper reported data about different countries, each country was separately evaluated. We classified the papers by adjustment for breast cancer risk (data, model and covariate used) and by

Table 2 Details of the adjustment for breast cancer risk

Paper	Model	Adjusted for...	Comments	Annual increase in the prescreening period
Peeters <i>et al.</i> (1989) ⁸	Mantel-Haenszel	Birth year	Birth-cohort specific incidence rates in a neighbouring unscreened area were used as reference	Not estimated
Paci <i>et al.</i> (2004) ⁹	Standardization	Age class (5-years)	The expected no. of cases in the absence of screening was estimated applying the age-specific prescreening incidence rates to the age distribution of the population during the study period	Not estimated
Zahl <i>et al.</i> (2004) ¹⁰	Poisson regression over all the study period	Age class (50–69, 70–74)	The annual percentage change was reported in the tables of the Results but it was not taken into account in estimating overdiagnosis	1.01 (0.99–1.02)
Zahl <i>et al.</i> (2004) ¹⁰	Poisson regression over all the study period	Age class (50–69, 70–74, 75–79)	The annual percentage change was reported in the tables of the Results but it was not taken into account in estimating overdiagnosis	1.008 (1.007–1.009)
Jonsson <i>et al.</i> (2005) ¹¹	Poisson regression on prescreening incidence (15 years of prescreening)	Year and county, stratified for four age class (40–49, 50–59, 60–69, 70–74)	The model included the year on the screening scale (from –15 to –1) instead of the calendar year	Not reported
Olsen <i>et al.</i> (2006) ¹²	Multistate modelling	Not necessary	The estimate of overdiagnosis was obtained fitting a model to the screening data. No adjustment for breast cancer risk is needed	Not estimated
Paci <i>et al.</i> (2006) ¹³	Poisson regression on prescreening incidence (6 years of prescreening)	Age (annual, 40–79), calendar year and area	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)	1.012 (1.008–1.016)
Waller <i>et al.</i> (2007) ¹⁴	Age-period-cohort model over all the study period	Age (annual, 20–89), calendar year, birth cohort, use of HRT and screening attendance	Estimates of the prevalence of HRT use were obtained from the General Practice Research Database	Not reported
Jørgensen and Gotzsche (2009) ¹⁵	Linear regression on prescreening incidence (1971–84)	Calendar year, stratified for three age class (30–49, 50–64, 65–74)	Linear regression instead of Poisson regression was fitted because the denominators for the rates were not available	Not reported
Jørgensen and Gotzsche (2009) ¹⁵	Linear regression on prescreening incidence (1971–85)	Calendar year, stratified for three age class (30–49, 50–69, 70–84)	Linear regression instead of Poisson regression was fitted because the denominators for the rates were not available	Not reported
Jørgensen and Gotzsche (2009) ¹⁵	Linear regression on prescreening incidence (1980–94)	Calendar year, stratified for three age class (40–49, 50–69, 70–79)	Linear regression instead of Poisson regression was fitted because the denominators for the rates were not available	not reported
Puliti <i>et al.</i> (2009) ¹⁶	Poisson regression on prescreening incidence (1986–90)	Age (annual, 40–79) and calendar year	Calendar year parameter was forced to the value previously estimated in central and northern Italy (ref 6). A sensitivity analysis assuming no trend was performed	1.012 (1.008–1.016)
Jørgensen <i>et al.</i> (2009) ¹⁷	Poisson regression on prescreening incidence (1971–1990)	Age class (5-years) and geographical area, stratified for screening age (50–69) and exceeded age (70–79)	The adjustment for prescreening geographical differences is not considered sufficient	1.0037 (1.0030–1.0045)
Duffy <i>et al.</i> (2010) ¹⁸	Poisson regression on prescreening incidence (1974–88)	Calendar year, stratified for age class (<45, 45–49, 50–64, 65–69, 70+)	Plus adjustment for non-linear trends (dividing the expected numbers by the relative excess for <45 years)	Not reported

Table 2 (Continued.)

Paper	Model	Adjusted for...	Comments	Annual increase in the prescreening period
Martinez-Alonso <i>et al.</i> (2010) ¹⁹	Age-cohort model over all the study period	Age (annual, 25–84), year of birth, fertility rate and use of mammography	This model was used to estimate the background incidence in the absence of screening assuming the use of mammography equal zero	Not estimated
de Gelder <i>et al.</i> (2011) ²⁰	Microsimulation modelling based on prescreening incidence data	Not necessary	Via micro simulation modelling the incidence without screening in a population aged 0–100 was predicted	Not estimated

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

Paper	Method used for adjustment for lead time	Comments
Paci <i>et al.</i> (2004) ⁹	Individually screen-detected cases were shifted forward (for age and calendar year) according to the distribution of lead time	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
Jonsson <i>et al.</i> 2005 ¹¹	A period equal to the 65% of lead time was added to age at diagnosis for all cases diagnosed in the screening period	(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
Olsen <i>et al.</i> (2006) ¹²	A multistate model was fitted to incidence of a cohort of women screened at least once	Assuming an exponential distribution of incidence for the non-progressive preclinical screen-detectable cancers and a test sensitivity of 100%
Paci <i>et al.</i> (2006) ¹³	Individually screen-detected cases were shifted forward (for age and calendar year) according to the distribution of lead time	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
Martinez-Alonso <i>et al.</i> (2010) ¹⁹	A probabilistic model (including background incidence, competitive risks, distribution of sojourn time, sensitivity and dissemination of screening) was fitted	Using the probabilistic model, the expected incidence due to lead time (i.e. incidence with screening assuming no overdiagnosis) was estimated

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^{9,10} used prescreening age distribution without considering temporal trend, seven^{11,13,15,16,18} used extrapolation of prescreening trends, two^{8,17} were geographically controlled, two^{14,19} used risk factor adjustment and two^{12,20} 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.*⁸ calculated overdiagnosis 12 years after the start of a pilot screening programme in Nijmegen, the Netherlands. The control population was represented by women aged ≥ 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 un-screened 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.*^{9,13} (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,⁹ 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

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

Paper	Incidence excess in the screened group		Compensatory drop in the older age groups			
	Age and period considered	Incidence excess (Obs/Exp)	Exceeded age for screening	% women who have had the opportunity to screen*	Mean (range) follow-up after screening stops*	Compensatory drop (Obs/Exp)
Zahl <i>et al.</i> (2004) ¹⁰	50–69 years in the 1998–99	+54%	70–74 years in the 2000	80%	2.5 years (1–4 years)	–11% [†]
Zahl <i>et al.</i> (2004) ¹⁰	50–69 years in the 1997–2000	+45%	(1) 70–74 years in 1997–2000 (2) 75–79 years in 1997–2000	(1) 100% (continuing screening) (2) 95%	(1) No follow-up (2) 4.2 years (1–10 years)	(1) Not observed (2) –12% [†]
Waller <i>et al.</i> (2007) ¹⁴	50–64 years in 1990–2001	+73% (first screen) +18–35% (subsequent)	65–67 years	100%	1–3 years	–12%
Jørgensen and Gotzsche (2009) ¹⁵	50–64 years in 1993–1999	+41%	68–70 years 71–73 years 65–74 years in 1993–1999	82%	4–6 years 7–9 years 4.4 years (1–10 years)	–8% –3% Not observed
Jørgensen and Gotzsche (2009) ¹⁵	50–69 years in 1998–2006	+35%	70–84 years in 1998–2006	94% (1/3 continuing screening)	2.3 years (0–15 years)	–10%
Jørgensen and Gotzsche (2009) ¹⁵	50–69 years in 2000–2006	+42%	70–79 years in 2000–2006	75%	3.9 years (1–10 years)	–15%
Puliti <i>et al.</i> (2009) ¹⁶	60–69 years in 1990–1999	+21%	70–83 years in 1991–2004 (belong to the cohort)	100%	4.7 years (1–14 years)	–13%
Jørgensen <i>et al.</i> (2009) ¹⁷	50–69 years in 1991–2003	+40%	70–79 years in 1998–2003	92%	4.6 years (1–10 years)	–10%
Duffy <i>et al.</i> (2010) ¹⁸	45–64 years in 1989–2003	+13%	65+ years in 1989–2003	60%	5.0 years (1–15 years)	–8.1%
De Gelder <i>et al.</i> (2011) ²⁰	0–74 years in 2006	+7%	75–100 years in 2006	79%	6.1 years (1–16 years)	–11.7%

*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 1995²³

[†]It should be noted that these compensatory drops (11% in Norway and 12% in Sweden) were not considered to estimate overdiagnosis

programme to the age distribution of the target population during the study period, without considering any temporal trend. In the later paper,¹³ 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.*¹¹ 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.*^{9,13} is different in several respects from that employed by Jonsson *et al.*¹¹ 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,²¹ it was stated that the results of the Swedish study reported in Jonsson *et al.*¹¹ were not explicitly attributed to overdiagnosis. Other potential factors were given (including changes in risk factor

Table 4 Details of the measure of overdiagnosis

Paper	Measure of overdiagnosis (OD)	Numerator	Denominator	Age range (years)
		Definition	Definition	
Peeters <i>et al.</i> (1989) ⁸	Excess incidence/Expected incidence	The difference between the observed incidence in screened area and the observed incidence in unscreened area adjusted for birth year	The observed incidence in the unscreened area adjusted for birth year	35+
Paci <i>et al.</i> (2004) ⁹	Adjusted excess incidence/Expected incidence	The difference between the post-screening incidence statistically adjusted for lead time and the prescreening incidence adjusted for age	Prescreening incidence adjusted for age	50–84
Zahl <i>et al.</i> (2004) ¹⁰	Excess incidence/Expected incidence	The difference between the post-screening incidence and the prescreening incidence adjusted for age	Prescreening incidence adjusted for age	50–69
Zahl <i>et al.</i> (2004) ¹⁰	Excess incidence/Expected incidence	The difference between the post-screening incidence and the prescreening incidence adjusted for age	Prescreening incidence adjusted for age	50–69
Jonsson <i>et al.</i> (2005) ¹¹	Adjusted excess incidence/Expected incidence	The difference between the post-screening incidence statistically adjusted for lead time and the prescreening incidence adjusted for age, temporal trend and area	Prescreening incidence adjusted for age, temporal trend and area	40–74
Olsen <i>et al.</i> (2006) ¹²	Adjusted excess incidence/Observed incidence among screened	The adjusted excess incidence was estimated by multistate modelling	Observed incidence among screened	50–69 followed for 6 years
Paci <i>et al.</i> (2006) ¹³	Adjusted excess incidence/Expected incidence	The difference between the post-screening incidence statistically adjusted for lead time and the prescreening incidence adjusted for age, temporal trend and area	Prescreening incidence adjusted for age, temporal trend and area	50–74
Waller <i>et al.</i> (2007) ¹⁴	Absolute increase of lifetime risk of breast cancer due to screening	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)	Not applicable (The measure was expressed as absolute increase)	–
Jørgensen and Gotzsche (2009) ¹⁵	Adjusted excess incidence/Expected incidence	The difference between the post-screening incidence (no compensatory drop was found) and the prescreening incidence adjusted for age and temporal trend	Prescreening incidence adjusted for age and temporal trend	50–64
Jørgensen and Gotzsche (2009) ¹⁵	Adjusted excess incidence/Expected incidence	The difference between the post-screening incidence corrected for compensatory drop and the prescreening incidence adjusted for age and temporal trend	Prescreening incidence adjusted for age and temporal trend	50–69
Jørgensen and Gotzsche (2009) ¹⁵	Adjusted excess incidence/Expected incidence	The difference between the post-screening incidence corrected for compensatory drop and the prescreening incidence adjusted for age and temporal trend	Prescreening incidence adjusted for age and temporal trend	50–69
Puliti <i>et al.</i> (2009) ¹⁶	Adjusted excess incidence/Expected incidence	The difference between the post-screening incidence corrected for compensatory drop and the prescreening incidence adjusted for age and temporal trend	Prescreening incidence adjusted for age and temporal trend	60–69 followed for 15 years
Jørgensen <i>et al.</i> (2009) ¹⁷	Adjusted excess incidence/Expected incidence	The difference between the post-screening incidence corrected for compensatory drop and the prescreening incidence adjusted for age	Prescreening incidence adjusted for age	50–69
Duffy <i>et al.</i> (2010) ¹⁸	No. cases overdiagnosed for 1000 women screened for 20 years	The difference between the post-screening incidence corrected for compensatory drop and the prescreening incidence adjusted for age and temporal trend	Person-years of screened women	–
Martinez-Alonso <i>et al.</i> (2010) ¹⁹	Adjusted excess incidence/Expected incidence	The difference between the observed incidence with screening and the expected incidence with screening due to lead time (assuming no overdiagnosis)	The expected incidence with screening due to lead time	40+
de Gelder <i>et al.</i> (2011) ²⁰	Adjusted excess incidence/Expected incidence	The difference between the predicted incidence with screening and predicted incidence without screening (as estimated by MISCAN)	Predicted incidence without screening (as estimated by MISCAN)	0–100

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.*¹² (Denmark) estimated the natural history of breast cancer by multistate modelling in a similar approach to that of Day and Walter.²² 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 ($0.0038 \times 112,860 = 429$), obtaining an estimate of 7.0% ($30/429$).

Martinez-Alonso *et al.*¹⁹ 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.*¹⁰ 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.*¹⁴ (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.*,²³ 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.²⁴ 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 Gotzsche¹⁵ 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 at 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 post-screening 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 Gøtzsche 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.

Puliti *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 rate 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 Gøtzsche¹⁵ 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

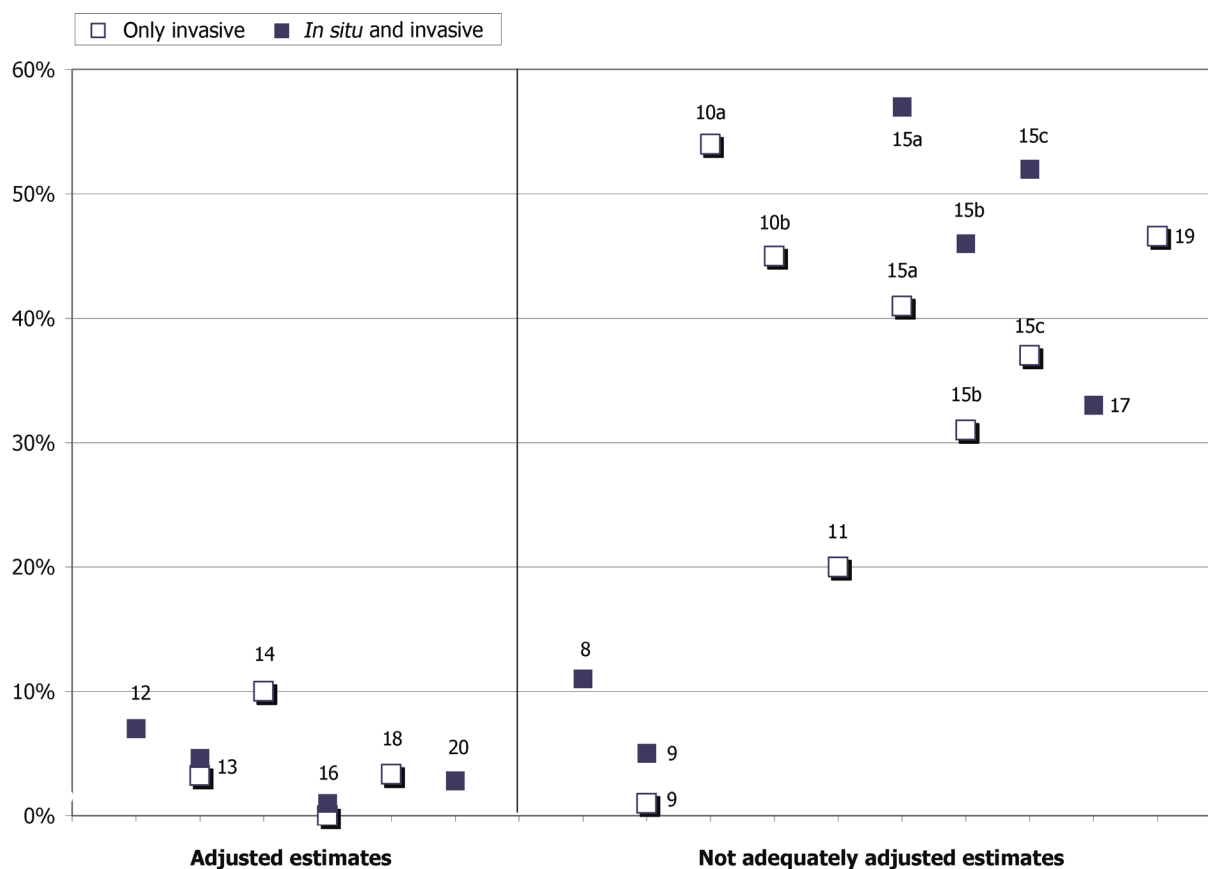


Figure 1 Overdiagnosis estimates classified according to the presence/absence of both the adjustments. The numbers indicate the related reference. Notes: (1) For the paper by Jonsson *et al.*,¹¹ we reported the pooled estimate for 40–74 years (20%) calculated by Jonsson himself. (2) For the paper by Martinez-Alonso *et al.*,¹³ we reported the estimate of the cohort of women born in 1950 considered by the authors themselves to be the best estimate (personal communication)

classified according to the presence/absence of both the adjustments. There is a clear difference between the two groups.

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.*¹⁸ 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.*² 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.

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 shown²⁰ 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.*,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.*¹² and by Waller *et al.*,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^{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.^{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.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.*,²³ is recommended.²⁴

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)¹, Armaroli P (I)², Ascunce N (E)³, Bisanti, L (I)⁴, Bellisario C (I)², Broeders M (NL)⁵, Cogo C (I)⁶, De Koning H (NL)⁷, Duffy SW (UK)⁸, Frigerio A (I)², Giordano L (I)², Hofvind S (N)⁹, Jonsson H (S)¹⁰, Lyng E (DK)¹¹, Massat N (UK)⁸, Miccinesi G (I)¹², Moss S (UK)⁸, Naldoni C (I)¹³, Njor S (DK)¹¹, Nyström I (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 de l'emploi 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

Donella Puliti, Statistician, Clinical and Descriptive Epidemiology Unit, ISPO – Cancer Research and Prevention Institute, Florence, Italy

Stephen W Duffy, Professor of Cancer Screening, Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, UK

Guido Miccinesi, Epidemiologist, Clinical and Descriptive Epidemiology Unit, ISPO – Cancer Research and Prevention Institute, Florence, Italy

Harry de Koning, Professor of Public Health and Screening Evaluation, Department of Public Health, Erasmus MC, Rotterdam, The Netherlands

Elsebeth Lynge, Professor of Epidemiology, Department of Public Health, University of Copenhagen, Denmark

Marco Zappa, Epidemiologist, Clinical and Descriptive Epidemiology Unit, ISPO – Cancer Research and Prevention Institute, Florence, Italy

Eugenio Paci, MD MPH, Director, Clinical and Descriptive Epidemiology Unit, ISPO – Cancer Research and Prevention Institute, Florence, Italy

the EUROSREEN Working Group (members listed at the end of the paper)

ACKNOWLEDGEMENTS

The financial support was provided by the National Monitoring Italian Centre (ONS) to host the EUROSREEN 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 EUROSREEN mortality working group in July 2011.

REFERENCES

- Paci E, Duffy S. Overdiagnosis and overtreatment of breast cancer: overdiagnosis and overtreatment in service screening. *Breast Cancer Res* 2005;**7**:266–70
- Kopans DB, Smith RA, Duffy SW. Mammography screening and 'overdiagnosis'. *Radiology* 2011;**260**:616–20
- Gotzsche PC, Hartling OJ, Nielsen M, Brodersen J, Jørgensen KJ. Breast screening: the facts – or maybe not. *BMJ* 2009;**338**:b86
- Welch HG, Black WC. Overdiagnosis in cancer. *J Natl Cancer Inst* 2010;**102**:605–13
- Biesheuvel C, Barratt A, Howard K, Houssami N, Irwig L. Effects of study methods and biases on estimates of invasive breast cancer overdiagnosis with mammography screening: a systematic review. *Lancet Oncol* 2007;**8**:1129–38
- Moss S. Overdiagnosis and overtreatment of breast cancer: overdiagnosis in randomised controlled trials of breast cancer screening. *Breast Cancer Res* 2005;**7**:230–4

- Zackrisson S, Andersson I, Janzon L, Manjer J, Garne JP. Rate of overdiagnosis of breast cancer 15 years after end of Malmö mammographic screening trial: follow-up study. *BMJ* 2006;**332**:689–92
- Peeters PH, Verbeek AL, Straatman H, et al. Evaluation of overdiagnosis of breast cancer in screening with mammography: results of the Nijmegen programme. *Int J Epidemiol* 1989;**18**:295–9
- Paci E, Warwick J, Falini P, Duffy SW. Overdiagnosis in screening: is the increase in breast cancer incidence rates a cause for concern? *J Med Screen* 2004;**11**:23–7
- Zahl PH, Strand BH, Maehlen J. Incidence of breast cancer in Norway and Sweden during introduction of nationwide screening: prospective cohort study. *BMJ* 2004;**328**:921–4
- Jonsson H, Johansson R, Lenner P. Increased incidence of invasive breast cancer after the introduction of service screening with mammography in Sweden. *Int J Cancer* 2005;**117**:842–7
- Olsen AH, Agbaje OF, Myles JP, Lynge E, Duffy SW. Overdiagnosis, sojourn time, and sensitivity in the Copenhagen mammography screening program. *Breast J* 2006;**12**:338–42
- Paci E, Miccinesi G, Puliti D, et al. Estimate of overdiagnosis of breast cancer due to mammography after adjustment for lead time. A service screening study in Italy. *Breast Cancer Res* 2006;**8**:R68
- Waller M, Moss S, Watson J, Moller H. The effect of mammographic screening and hormone replacement therapy use on breast cancer incidence in England and Wales. *Cancer Epidemiol Biomarkers Prev* 2007;**16**:2257–61
- Jørgensen KJ, Gotzsche P. Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. *BMJ* 2009;**339**:b2587
- Puliti D, Zappa M, Miccinesi G, Falini P, Crocetti E, Paci E. An estimate of overdiagnosis 15 years after the start of mammographic screening in Florence. *Eur J Cancer* 2009;**45**:3166–71
- Jørgensen KJ, Zahl PH, Gotzsche P. Overdiagnosis in organised mammography screening in Denmark. A comparative study. *BMC Women's Health* 2009;**9**:36
- Duffy SW, Tabar L, Olsen AH, et al. Absolute numbers of lives saved and overdiagnosis in breast cancer screening, from a randomized trial and from the Breast Screening Programme in England. *J Med Screen* 2010;**17**:25–30
- Martinez-Alonso M, Vilapinyo E, Marcos-Gragera R, Rue M. Breast cancer incidence and overdiagnosis in Catalonia (Spain). *Breast Cancer Res* 2010;**12**:R58
- de Gelder R, Heijnsdijk EA, van Ravestein NT, Fracheboud J, Draisma G, de Koning HJ. Interpreting overdiagnosis estimates in population-based mammography screening. *Epidemiol Rev* 2011;**33**:111–21
- Duffy SW, Lynge E, Jonsson H, Ayyaz S, Olsen AH. Complexities in the estimation of overdiagnosis in breast cancer screening. *Br J Cancer* 2008;**99**:1176–8
- Day NE, Walter SD. Simplified models of screening for chronic disease: estimation procedures from mass screening programmes. *Biometrics* 1984;**40**:1–14
- Moller B, Weedon-Fekjaer H, Hakulinen T, et al. The influence of mammographic screening on national trends in breast cancer incidence. *Eur J Cancer Prev* 2005;**14**:117–28
- Duffy SW, Jonsson H, Agbaje OF, Pashayan N, Gabe R. Avoiding bias from aggregate measures of exposure. *J Epidemiol Community Health* 2007;**61**:461–3
- Andreasen AH, Andersen KW, Madsen M, Mouridsen H, Olesen KP, Lynge E. Regional trends in breast cancer incidence and mortality in Denmark prior to mammographic screening. *Br J Cancer* 1994;**70**:133–7
- Puliti D, Miccinesi G, Paci E. Overdiagnosis in breast cancer: design and methods of estimate in observational studies. *Prev Med* 2011;**53**:131–3.
- Boer R, Warmerdam P, de Koning H, van Oortmarsen G. Extra incidence caused by mammographic screening. *Lancet* 1994;**343**:979
- Walter SD, Day NE. Estimation of the duration of a pre-clinical disease state using screening data. *Am J Epidemiol* 1983;**118**:865–86

APPENDIX

Search Strategy

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

- overdiagnosis
- mammography screening

(3) (#1) AND #2

This search strategy retrieved a total of 99 papers.

(4) 'Mass Screening'[Mesh]

(5) 'Mammography'[Mesh]

(6) 'Breast Neoplasms'[Mesh]

(7) 'Diagnostic Errors'[Mesh]

(8) 'False Positive Reactions'[Mesh]

(9) 'Reproducibility of Results'[Mesh]

(10) 'Sensitivity and Specificity'[Mesh]

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

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

(13) (#12) AND #11

(14) (#7) OR #8

(15) (#12) AND #14

(16) (#14) OR #1

(17) (#12) AND #16

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

(19) **(#18) NOT #3**

This search strategy retrieved a total of 382 papers.

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

(21) **(#20) NOT #3**

This search strategy retrieved a total of 1040 papers.

(22) ((#4 AND #6) AND #2

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

(24) **(#22) AND #23**

This search strategy retrieved a total of 1168 papers.

(25) **(#1) AND #4) AND #6**

This search strategy retrieved a total of 83 papers.

(26) harm and benefit

(27) **(#26) AND #2**

This search strategy retrieved a total of 37 papers.

(28) advantages and disadvantages

(29) **(#28) AND #2**

This search strategy retrieved a total of 45 papers.

(30) 'Incidence'[Mesh]

(31) **(#30) AND #4) AND #6**

This search strategy retrieved a total of 596 papers.

Publications of authors expert in the field:

(32) Duffy S[Author] AND #1

This search strategy retrieved a total of 15 papers.

(33) Paci E[Author] AND #1

This search strategy retrieved a total of 6 papers.

(34) Lynge E[Author] AND #1

This search strategy retrieved 8 articles.

(35) Zahl PH[Author] AND #1

This search strategy retrieved 8 articles.

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

(36) Paci E, Duffy S. Overdiagnosis and overtreatment of breast cancer: overdiagnosis and overtreatment in service screening. *Breast Cancer Res* 2005;**7**(6): 266–70

This function retrieved a total of 240 papers.

(37) Paap E, Verbeek AL, Puliti D, Paci E, Broeders MJ. Breast cancer screening case-control study design: impact on breast cancer mortality. *Ann Oncol* 5 October 2010

This function retrieved a total of 104 papers.

(38) Welch HG, Black WC. Overdiagnosis in cancer. *Natl Cancer Inst* 5 May 2010;**102**(9):605–13.

This function retrieved a total of 95 papers.

(39) Welch HG. Screening mammography – a long run for a short slide? *N Engl J Med* 23 September 2010; 363(13):1276–8

This function retrieved a total of 106 papers.

(40) Puliti D, Zappa M, Miccinesi G, Falini P, Crocetti E, Paci E. An estimate of overdiagnosis 15 years after the start of mammographic screening. *Florence. Eur J Cancer* December 2009;**45**(18):3166–71

This function retrieved a total of 119 papers.

(41) Jørgensen KJ, Zahl PH, Gøtzsche PC. Breast cancer mortality in organised mammography screening in Denmark: comparative study. *BMJ* 23 March 2010;**340**:c1241

This function retrieved a total of 113 papers.

(42) 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.

(43) Kalager M, Zelen M, Langmark F, Adami HO. Effect of screening mammography on breast-cancer mortality in Norway. *N Engl J Med* 23 September 2010;**363**(13):1203–10.

This function retrieved a total of 115 papers.

(44) Morrell S, Barratt A, Irwig L, Howard K, Biesheuvel C, Armstrong B. Estimates of overdiagnosis of invasive breast cancer associated with screening mammography. *Cancer Causes Control* February 2010;**21**(2): 275–82.

This function retrieved a total of 174 papers.

(45) Esserman L, Thompson I. Solving the overdiagnosis dilemma. *J Natl Cancer Inst* 5 May 2010;**102**(9):582–3

This function retrieved a total of 124 papers.

(46) Seppänen J, Heinävaara S, Anttila A, Sarkeala T, Virkkunen H, Hakulinen T. Effects of different phases of an invitational screening program on breast cancer incidence. *Int J Cancer* 15 August 2006;**119**(4):920–4

This function retrieved a total of 150 papers.

References from the following published articles:

- Evans A, Cornford E, James J. Breast screening overdiagnosis. Stop treating indolent lesions. *BMJ* 11 August 2009; 339
- Welch HG. Overdiagnosis and mammography screening. *BMJ* 9 July 2009;339
- Esserman L, Thompson I. Solving the overdiagnosis dilemma. *J Natl Cancer Inst* 5 May 2010 ;102(9):582–3
- Newman DH. Screening for breast and prostate cancers: moving toward transparency. *J Natl Cancer Inst* 21 July 2010;102(14):1008–11

- 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.osservatorionazionale screening.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.