Interpreting Overdiagnosis Estimates in Population-based Mammography Screening

Rianne de Gelder*, Eveline A. M. Heijnsdijk, Nicolien T. van Ravesteyn, Jacques Fracheboud, Gerrit Draisma, and Harry J. de Koning

* Correspondence to Rianne de Gelder, Department of Public Health, Erasmus MC, Room AE-137, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands (e-mail: r.degelder@erasmusmc.nl).

Estimates of overdiagnosis in mammography screening range from 1% to 54%. This review explains such variations using gradual implementation of mammography screening in the Netherlands as an example. Breast cancer incidence without screening was predicted with a micro-simulation model. Observed breast cancer incidence (including ductal carcinoma in situ and invasive breast cancer) was modeled and compared with predicted incidence without screening during various phases of screening program implementation. Overdiagnosis was calculated as the difference between the modeled number of breast cancers with and the predicted number of breast cancers without screening. Estimating overdiagnosis annually between 1990 and 2006 illustrated the importance of the time at which overdiagnosis is measured. Overdiagnosis was also calculated using several estimators identified from the literature. The estimated overdiagnosis rate peaked during the implementation phase of screening, at 11.4% of all predicted cancers in women aged 0–100 years in the absence of screening. At steady-state screening, in 2006, this estimate had decreased to 2.8%. When different estimators were used, the overdiagnosis rate in 2006 ranged from 3.6% (screening age or older) to 9.7% (screening age only). The authors concluded that the estimated overdiagnosis rate in 2006 could vary by a factor of 3.5 when different denominators were used. Calculations based on earlier screening program phases may overestimate overdiagnosis by a factor 4. Sufficient follow-up and agreement regarding the chosen estimator are needed to obtain reliable estimates.

breast neoplasms; early detection of cancer; incidence; mammography; mass screening; overdiagnosis; risk

Abbreviation: DCIS, ductal carcinoma in situ.

INTRODUCTION

Mammography screening has been shown to be effective in reducing breast cancer mortality (1–4), but the magnitude of the harms of screening is less well established. One of the harms of screening is overdiagnosis: detection of breast cancers that would not have become symptomatic during a woman's lifetime if no screening had taken place.

Screening is expected to increase the observed incidence of breast cancer among women in the targeted age group, partly because of overdiagnosis but also because of the advanced diagnosis of breast cancer. At prevalence screening, mammography may detect breast cancers from a pool of preclinical tumors that exist in a population, which increases the observed incidence. At subsequent screens, future incidence trends—breast cancer incidence in industrialized countries increases by calendar time (4)—are brought for-

ward in time, which may also lead to an excess of breast cancers compared with a situation without screening.

Theoretically, those tumors for which the diagnosis is advanced by screening will not be diagnosed when they would have been if no screening had taken place. The period during which the diagnosis is advanced is called "lead time." When the lead time has elapsed, the incidence among previously screened women is expected to fall to a level below that predicted without screening ("deficit incidence"). The deficit in incidence in the previously screened age group is expected to balance out the excess in incidence in the screening ages. In practice, this is not entirely the case (5–7) and is referred to as overdiagnosis.

The frequency at which overdiagnosis occurs is a topic of strong debate. In a meta-analysis of overdiagnosis in randomized breast cancer screening trials and various population-based screening programs, Biesheuvel et al. (8)

found that estimates ranged between −4% and 54% of all expected cancers (including invasive breast cancers only). More recent analyses estimated the rate of overdiagnosis to be 52%–54% of all expected cancers without screening in women of the screening age, meaning that 1 of 3 cancers in a screened population is overdiagnosed (5, 9). Modeling studies, on the contrary, estimated overdiagnosis to be between 1% of all diagnosed breast cancers in a screened population (6) and 3% of all predicted breast cancers in the total population (10).

The question thus arises regarding why overdiagnosis estimates differ to such an extent between those studies. In this analysis, we discuss differences between key studies of overdiagnosis. Using the gradual implementation of nationwide breast cancer screening of more than 1 million Dutch women as an example, we focus on the importance of the time at which overdiagnosis is measured. To enable a comparison between various estimates of overdiagnosis, we calculate rates using several different denominators.

MATERIALS AND METHODS

MIcro-simulation SCreening ANalysis model (MISCAN)

Biennial mammography screening in the Netherlands started in 1990 and was gradually implemented in the whole country between 1990 and 1997 ("implementation phase"), targeting women in the age group 49–69 years. Between 1998 and 2001, the screening program was extended to women aged 49–74 years ("extension phase"). A "steady-state phase" of screening was reached in 2002, when the number of first screening examinations and subsequent examinations with an interval of more than 30 months from the previous screening remained stable. Annually, more than 1 million women are invited to participate in the program; 82% of the invited population attends screening.

The implementation, extension, and steady-state phases of the Dutch screening program were modeled in MISCAN (11, 12), designed to assess the effects of screening on a population. The model consisted of 2 parts: a part in which 1) the individual life histories of women in a nonscreened population were simulated, and 2) a part in which a screening program was modeled and the influence of mammography screening on these life histories was determined. We modeled the Dutch female population aged 0–100 years in 1989. In the model, some of these women may develop preclinical invasive breast cancer during their lives, which may or may not be preceded by screen-detectable ductal carcinoma in situ (DCIS). In the absence of screening, preclinical screen-detectable DCIS may progress to invasive cancer, become clinically diagnosed, or regress. Preclinical invasive tumors may grow into a successively larger preclinical stage of disease, as a Markov-like stage transition process. They also may become symptomatic and consequently diagnosed. If screening takes place, preclinical lesions can also become screen detected, depending on their size and the sensitivity of the test.

The rates at which these transitions occur, the mean duration of preclinical DCIS and invasive cancer, and the sensitivity of screening mammography were estimated by

using data from the Comprehensive Cancer Centers (13) and screening organizations in the Netherlands (14). These data included the observed age-specific incidence in the Netherlands between 1990 and 2006; the age-specific and stage-specific incidence of clinically diagnosed and screen-detected cancer; the age-, stage-, and screening-round–specific cancer detection rates; and interval cancer rates since the start of screening. By minimizing the deviance between observed and modeled breast cancer incidence, screen detection, and interval cancer rates, the optimal model parameters were chosen. A chi-square test was used to test goodness of fit. With the model, the observed incidence in the Netherlands and various other countries could be reproduced reasonably (15–17). We assumed that no mammography screening took place outside the organized program because screening participation is high, especially among women who have previously attended (82% in the target population, 95% of previous attendees) (14). Furthermore, a survey of women older than the screening age showed no evidence for opportunistic screening (18).

Model parameters

In the best-fitting model, we estimated the duration of preclinical DCIS to be Weibull distributed, with a mean of 2.6 years at all ages. The duration of preclinical invasive cancer has an exponential distribution with a mean estimated to increase by age, from 1.0 year at age 20 to 3.9 years from age 65 onward. The sensitivity of mammography was 72% for DCIS, 47% for stage T1a, 62% for stage T1b, 90% for stage T1c, and 95% for stage T2+, and it was assumed not to change over time. We further estimated that 18% of all tumors had a screen-detectable DCIS stage, of which 11% progressed to invasive breast cancer, 5% was clinically diagnosed, and 2% regressed.

Overdiagnosis calculations

To calculate the rate of overdiagnosis in the Netherlands, we modeled breast cancer incidence in the presence of screening. The model that fitted best to the observations was then compared with predicted breast cancer incidence in the absence of screening. With this approach, screened and nonscreened populations were exactly the same, with a similar background risk of developing breast cancer. Both DCIS and invasive cancer were included in the overdiagnosis estimate, because, in the model, overdiagnosis can occur when

1. preclinical DCIS, detected by screening, would have regressed if no screening had taken place (Figure 1A);
2. preclinical DCIS, detected by screening, would not have progressed to invasive cancer during a woman's lifetime if no screening had taken place (Figure 1B);
3. preclinical DCIS, detected by screening, would have progressed to invasive cancer but would not have become symptomatic during a woman's lifetime if no screening had taken place (Figure 1C); and
4. preclinical invasive breast cancer, detected by screening, would not have become symptomatic during a woman's lifetime if no screening had taken place (Figure 1D).
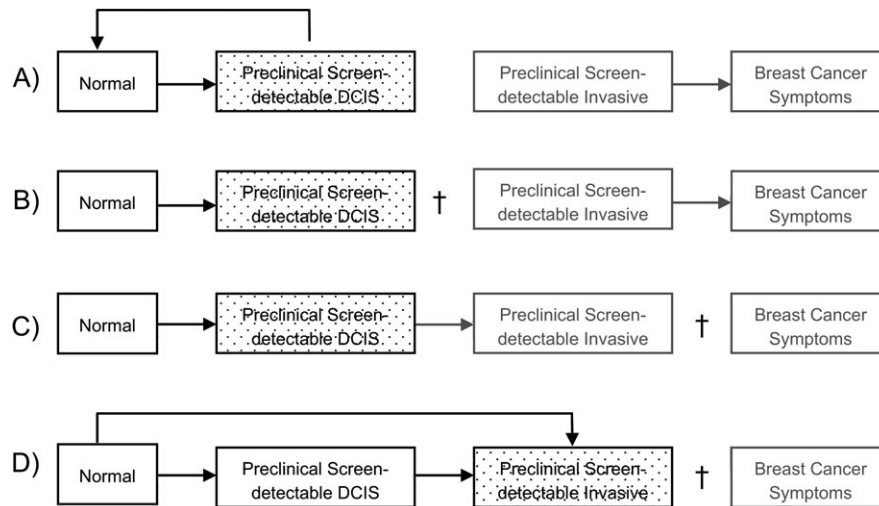
Figure 1. Stages at which overdiagnosis can occur: A) When preclinical ductal carcinoma in situ (DCIS), detected by screening, would have regressed if no screening had taken place; B) when preclinical DCIS, detected by screening, would not have progressed to invasive cancer during a woman's lifetime if no screening had taken place; C) when preclinical DCIS, detected by screening, would have progressed to invasive cancer but would not have become symptomatic during a woman's lifetime if no screening had taken place; and D) when preclinical invasive breast cancer, detected by screening, would not have become symptomatic during a woman's lifetime if no screening had taken place. Dot-filled boxes; stage at which a breast cancer is screen detected; grey-shaded boxes: stages of the natural history of the tumor averted by screen detection; crosses: death from causes other than breast cancer.

The estimator for the overdiagnosis rate was $(E - D)/T_{0, \text{ age 0–100 years}}$. E represents the number of excess breast cancers in women of screening age, calculated as the difference in the modeled number of breast cancers with and the predicted number of cancers without screening. D is the number of deficit breast cancers in the age groups exceeding the screening limit, calculated as the difference in the predicted number of breast cancers without and the modeled number of cancers with screening. $T_{0, \text{ age 0–100 years}}$ represents the total number of breast cancers predicted in a population aged 0–100 years without screening.

To illustrate the extent to which overdiagnosis estimates are influenced by the denominator used, various estimators are applied to the modeled and predicted breast cancer incidence in the Netherlands. We searched the PubMed literature to identify alternative estimators to calculate overdiagnosis. With the query "('breast neoplasms'[MeSH Terms] OR ('breast'[All Fields] AND 'cancer'[All Fields]) OR 'breast cancer'[All Fields]) AND ('overdiagnosis'[All Fields] OR 'over-diagnosis'[All Fields] OR 'overdetection'[All Fields] OR 'over-detection'[All Fields])," a total of 158 titles were obtained. Only primary research or review articles in English that gave explicit estimates of overdiagnosis in breast cancer screening trials and population-based mammography screening were considered relevant. Using these criteria, we included a total of 15 papers. On the basis of the literature references in these articles, 1 other paper was also included. Data on the denominator used to define the population at risk, the time period of screening, and the length of follow-up after screening ended were extracted from each study. An overview of the 16 studies obtained, with their estimates of overdiag-

nosis, is presented in Table 1. The studies were grouped by the estimator used to calculate overdiagnosis:

1. $(E - D)/T_{0, \text{ age 0–100 years}}$, which is the relative increase in breast cancers due to overdiagnosis $(E - D)$ compared with the predicted number of breast cancers in the female population aged 0–100 years in a situation without screening. This estimator was used by de Koning et al. (10), who estimated the overdiagnosis rate to be 3%.

2. $(E - D)/T_{0, \text{ screening age and older}}$, which is the relative increase in breast cancers due to overdiagnosis $(E - D)$ compared with the predicted number of breast cancers in women of the screening age and older in a situation without screening. Three previous studies used this estimator, with overdiagnosis estimates ranging between 1% and 30.5% (7, 19, 20).

3. $(E - D)/T_{0, \text{ screening age}}$, which is the relative increase in breast cancers due to overdiagnosis compared with the predicted number of breast cancers in women of the screening age in a situation without screening. This method was used in 3 studies (5, 21–23), with overdiagnosis estimates varying between 4.6% and 52%.

4. $(E - D)/T_{1, \text{ screening age}}$, which is the fraction of overdiagnosed cancers of all diagnosed breast cancers in women of the screening age in a situation with screening. The 3 studies that used this estimator assessed overdiagnosis to be 1%–12% (6, 24, 25).

5. $(E - D)/\text{SD}$, which is the fraction of all screen-detected (SD) cancers that is overdiagnosed. Welch and Black (26) recalculated the results of the Swedish Malmö trial (7) with this denominator and estimated the overdiagnosis rate to be 24%.

Table 1. Estimators for Overdiagnosis and Follow-up Time to Correct for Lead Time, as Reported in the Literature

Estimator	Method Used: First Author, Year (Reference No.)	Follow-up Allowed to Correct for Lead Time	Overdiagnosis Estimate
1. $(E - D)/T_{0, \text{age } 0-100 \text{ years}}$	de Koning, 2006 (10)	Modeled follow-up during remaining lifetime, in the steady-state phase of the screening program	3%
2. $(E - D)/T_{0, \text{screening age and older}}$	Moss, 2005 (19)	5–13 years of follow-up after randomization	−5.8% to 30.5%
	Zackrisson, 2006 (7)	15 years of follow-up after the trial ended	10%
	Puliti, 2009 (20)	5–10 years of follow-up past the screening age	1% to 13%
3. $(E - D)/T_{0, \text{screening age}}$	Paci, 2006 (21)	Modeled follow-up during remaining lifetime, in the first 5 years of the screening program	4.6%
	Jorgensen, 2009 (4 countries) (5)	Follow-up of 7–9 years after full implementation of screening or 10–11 years after the program started	52%
	Jorgensen, 2009 (Denmark) (5)	Follow-up of 2–10 years after full implementation of the program	33%
4. $(E - D)/T_{1, \text{screening age}}$	Duffy, 2005 (6)	Modeled follow-up during remaining lifetime, at the end of the screening trial	1% to 2%
	Olsen, 2006 (24)	Modeled follow-up during remaining lifetime, in the first 2 screening rounds	4.8%
	Duffy, 2010 (Sweden) (25)	Modeled follow-up during remaining lifetime, at the end of the screening trial	12%
5. $(E - D)/SD$	Welch, 2010 (26)	15 years of follow-up after the trial ended	24%
6. $T_{1, \text{screening age}}/T_{0, \text{screening age}}$	Zahl, 2004 (27)	1–4 years of follow-up after full implementation of the screening program	45% to 54% (excluding DCIS)
	Jonsson, 2005 (28)	7–15 years of follow-up since screening started	−4% to 54% (excluding DCIS)
	Morrell, 2010 (9)	4–6 years of follow-up after full implementation of the program	30% to 42% (excluding DCIS)
7. $T_{1, \text{screening age}}/T_{1, \text{screening age, corr}}$	Martinez-Alonso, 2010 (29)	Modeled follow-up during remaining lifetime, since screening started	0.4% to 46.6%

Abbreviations: D, number of deficit breast cancers in the age groups exceeding the screening limit, calculated as the difference in the number of breast cancers without and with screening; DCIS, ductal carcinoma in situ; E, number of excess breast cancers in the screening ages, calculated as the difference in the number of breast cancers with and without screening; SD, number of screen-detected cancers; T_0, predicted number of breast cancers in the absence of screening; T_1, modeled total number of breast cancers in the presence of screening; $T_{1, \text{corr}}$, total number of breast cancers in the presence of screening minus the number of overdiagnosed cancers.

6. $T_{1, \text{screening age}}/T_{0, \text{screening age}}$, which is the relative risk of breast cancer for women of the screening age in a situation with screening compared with the predicted number of breast cancers in women of the same age in a situation without screening. The estimator can be corrected for lead time, for instance, by shifting the predicted incidence without screening forward in time. Three studies used this method (9, 27, 28); their overdiagnosis estimates ranged between −4% and 54%.

7. $T_{1, \text{screening age}}/(T_{1, \text{screening age, corrected}})$, which is the relative risk of breast cancer for women of the screening age in a situation with screening compared with the predicted number of tumors in a situation with screening if no overdiagnosis would take place ($T_{1, \text{screening age, corrected}}$). This method was used by Martinez-Alonso et al. (29), who estimated the overdiagnosis rate to range between 0.4% and 46.6%.

Overdiagnosis was calculated by applying estimators 1–6 to the modeled and predicted numbers of breast cancers with and without screening in the Netherlands. Doing so demonstrates the impact of using different denominators on the estimated overdiagnosis rate. The overdiagnosis rate using estimator 6 ($T_{1, \text{screening age}}/T_{0, \text{screening age}}$) was calculated without a correction for lead time. A lead-time correction—for instance, by shifting the expected incidence without screening 2.5 years forward in age (comparable to the studies by Morrell et al. (9) and Jonsson et al. (28)—should result in an estimate in between those of estimators 3 and 6. Estimator 7 was not used, because it is not possible in MIS-CAN to model the incidence of breast cancer without assuming some degree of overdiagnosis. The outcomes were compared with the overdiagnosis rate obtained by using estimator 1 $(E - D)/T_{0, \text{age } 0-100 \text{ years}}$.

To illustrate the importance of the time at which overdiagnosis is estimated, the rate was calculated for each year between 1990 and 2006, during the implementation, extension, and steady-state phases of the screening program. Only in a steady state will the estimators provide an unbiased estimate of overdiagnosis.

RESULTS

From the moment that screening started in the Netherlands, observed breast cancer incidence among women of the screening ages increased (Figure 2 A–J). Related to the growing number of women screened and the relatively high

proportion of prevalence screens during the implementation phase of the program, the difference between the observed incidence rate of women of the screening ages and the predicted incidence in the absence of screening ("excess incidence") increased. At the end of the implementation phase (1996–1997), the invitation rate in the population no longer increased and excess incidence remained stable. Because part of the women aged 50–54 years reached the lower age limit for screening and had a prevalence screening, their incidence was higher than that of women aged 55–59 years, of whom the majority was invited for a subsequent screening round at this time.

In 1998, the upper age limit for the screening program was extended to women aged 70–74 years. At the peak of this extension phase, in 1999, the number of invited women and screening examinations with an interval of more than 2.5 years from the previous screening examination rose strongly, resulting in a higher detection rate. Consequently, excess breast cancer incidence among women of the screening ages also increased sharply (Figure 2F). The excess dropped again when all women aged 70–74 years had been reinvited to screening at least once (in 2002, Figure 2H) and a "steady-state" phase of the screening program was reached. From 2002 onward, the excess incidence among women of the screening ages remained fairly constant.

In the age groups that passed the screening age ($\geq$70 years between 1990 and 1997, $\geq$75 years from 1998 onward), the observed breast cancer incidence dropped to a level lower than the predicted incidence without screening (Figure 2). Because of lead time, generally estimated to be between 2 and 4 years (20, 30), the drop in incidence among women no longer screened is predicted to occur 2–4 years later than the increase in the screening ages, when all tumors would have been clinically diagnosed if no screening had taken place. Indeed, from 1994 onward, a deficit in breast cancer incidence was observed. From the moment that the majority of women were invited for a subsequent screening, in 1996, the deficit reached its maximum. The deficit in the incidence rate almost disappeared in the year the screening program was extended to include women aged 70–74 years. This extension phase lasted until 2001; the deficit in incidence among women aged 75 years or older was expected to be observed between 2003 and 2005. Indeed, the deficit increased during these years.

Our overdiagnosis estimates were based on the modeled incidence of breast cancer and the predicted incidence without screening. Overall, the model reproduced the observed incidence reasonably well. Between 1990 and 1993, however, the simulated incidence among women of the screening ages (50–69 years) was higher than observed, whereas, between 2001 and 2006, the modeled incidence was lower. When the modeled breast cancer incidence in a screening situation was compared with the predicted incidence without screening, the estimated overdiagnosis rate in the total population during the implementation phase of screening increased from 1.0% of all predicted breast cancers in 1990 to 11.4% in 1993 (Table 2). In 1993, the modeled excess in breast cancers peaked (17.1% of all predicted cancers in women aged 50–69 years), while the modeled deficit in incidence among women no longer screened was

0.8% of all predicted cancers in that age group. The estimate of overdiagnosis decreased the more women had subsequent screens, to 5.6% in 1997.

During the extension phase, the overdiagnosis estimate increased to 10.0% in 1999, after which it decreased to 4.7% in 2001. During the steady-state phase of screening, the estimate first increased to 4.9% in 2003 but then dropped to 2.8% of all predicted breast cancers in 2006. In 2006, the excess of breast cancers in the age group was 7.0%; the deficit was 11.7% (Table 2). Most of the deficit was expected directly when screening ceased: in the age group 70–74 years before 1998 and in the age group 75–79 years from 1998 onward. In 2006, a small deficit was also predicted among women aged 80–84 years.

Depending on the denominator used to define the population at risk, the overdiagnosis estimate at steady-state screening may increase to 8.9% if the rate is calculated as a fraction of all screen-detected cancers (Table 3, estimator 5, 2006). This rate is 3.2 times higher than the estimate that uses all predicted breast cancers in women aged 0–100 years in the denominator (Table 3, estimator 1, 2006) but has the same numerator. If calculated as a fraction of all diagnosed tumors among women of the screening age in a screening situation (Table 3, estimator 4, 2006), the estimate would be 4.6%. The estimated rates of overdiagnosis calculated as a relative increase among women of the screening age and older (Table 3, estimator 2, 2006) or women of the screening age only (Table 3, estimator 3, 2006) were 3.6% and 5.0%, respectively. Without an adjustment for lead time, the overdiagnosis rate calculated for women of the screening age only would be 9.7%: 3.5 times higher than the baseline estimate (Table 3, estimator 6, 2006). Overdiagnosis also depended on the year it was measured. Calculations based on years in which a screening program was not yet fully implemented were 4 times higher than estimates based on steady screening (Table 3, estimator 1, 1993 vs. 2006). The estimated overdiagnosis rate by year of measurement and by estimator is shown in Table 3.

DISCUSSION

The estimated overdiagnosis rate peaked during the implementation phase of screening at 11.4% of all predicted cancers in women aged 0–100 years in the absence of screening. Five years after implementation was completed, in 2006, this estimate had decreased to 2.8%. If different estimators were used, the overdiagnosis rate in 2006 would range between 3.6% (screening age and older) and 9.7% (screening age only). The estimate of overdiagnosis is thus strongly dependent on the time it was calculated and the denominator used to define the population at risk.

Our findings seem to strongly differ from those in some recent publications, with estimated overdiagnosis rates up to approximately 50% (5, 8, 9, 27). This paper may perhaps not resolve the controversy, but it does explain why reported epidemiologic estimates may differ to such an extent. Using gradual implementation of nationwide breast cancer screening of more than 1 million Dutch women, we illustrated that a steady-state screening situation and sufficient follow-up to
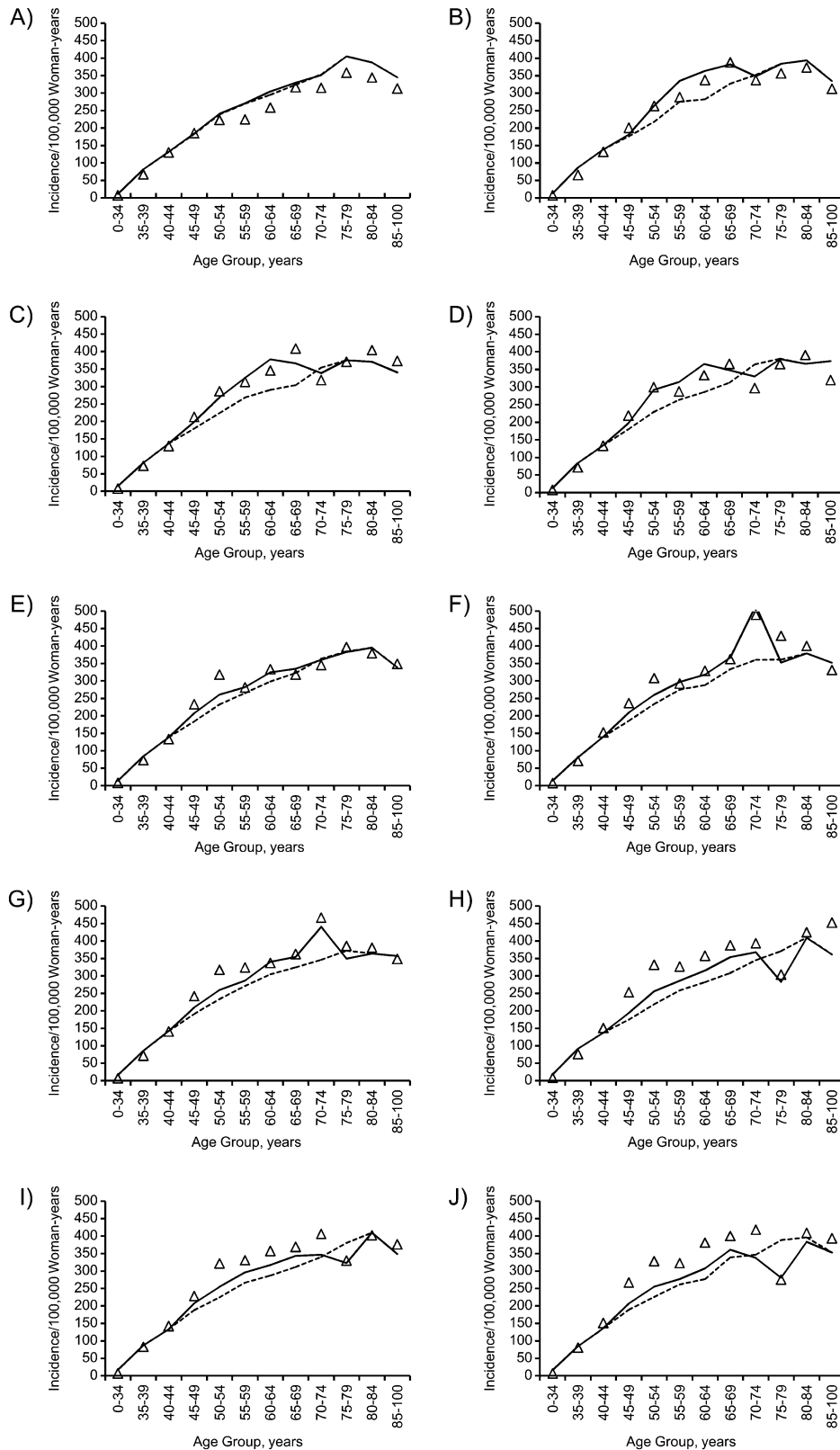
Figure 2. Observed and modeled breast cancer incidence per 100,000 woman-years in the presence and absence of screening between 1990 and 2006 (values after years indicate percentage of the target population aged 49–69 years invited, fraction of prevalent screenings). A) 1990: 9.2%, 74%; B) 1992: 47.4%, 77%; C) 1994: 74.3%, 49%; D) 1996: 92.0%, 39%; E) 1998: 80.8%, 20%; F) 1999: 91.8%, 19%; G) 2000: 94.4%, 18%; H) 2002: 96.1%, 14%; I) 2004: 95.8%, 14%; J) 2006: 92.2%, 13%. Solid lines, modeled with screening; dashed lines, modeled without screening; triangles, observed.

Table 2. Predicted Excess and Deficit in Breast Cancers and Overdiagnosis in the Netherlands[a]

Phase and Years	$T_{0,\ age\ 0-69/74\ years}$	$T_{1,\ age\ 0-69/74\ years}$	$E_{age\ 0-69/74\ years}$, %	$T_{0,\ age\ 69/74-100\ years}$	$T_{1,\ age\ 69/74-100\ years}$	$D_{age\ 69/74-100\ years}$, %	$E_{age\ 0-69/74\ years} - D_{age\ 69/74-100\ years}$	$(E-D)/T_{0,\ age\ 0-100\ years}$, %
Implementation phase								
1990–1991	15,237	15,481	1.6	7,207	7,197	0.1	234	1.0
1991–1992	15,646	17,065	9.1	7,201	7,184	0.2	1,402	6.1
1992–1993	15,606	17,719	13.5	7,240	7,214	0.4	2,087	9.1
1993–1994	15,695	18,381	17.1	7,458	7,400	0.8	2,628	11.4
1994–1995	16,039	18,490	15.3	7,499	7,405	1.3	2,357	10.0
1995–1996	16,149	18,550	14.9	7,821	7,669	1.9	2,249	9.4
1996–1997	16,235	18,608	14.6	7,877	7,628	3.2	2,124	8.8
1997–1998	16,646	18,291	9.9	7,958	7,686	3.4	1,373	5.6
Extension phase								
1998–1999	19,506	20,746	6.4	5,404	5,392	0.2	1,228	4.9
1999–2000	19,779	22,368	13.1	5,488	5,433	1.0	2,534	10.0
2000–2001	20,043	22,108	10.3	5,675	5,517	2.8	1,907	7.4
2001–2002	20,375	21,892	7.4	5,841	5,560	4.8	1,236	4.7
Steady-state phase								
2002–2003	20,371	21,961	7.8	5,892	5,538	6.0	1,236	4.7
2003–2004	20,601	22,336	8.4	5,965	5,533	7.2	1,303	4.9
2004–2005	20,471	22,127	8.1	5,908	5,377	9.0	1,125	4.3
2005–2006	20,984	22,741	8.4	5,857	5,288	9.7	1,188	4.4
2006–2007	21,087	22,569	7.0	6,136	5,421	11.7	767	2.8

[a] The percentage of excess (E) breast cancers in the age group 0–69/74 years was calculated as $(T_{1,\ age\ 0-69/74\ years} - T_{0,\ age\ 0-69/74\ years})/T_{0,\ age\ 0-69/74\ years}$. T_1, modeled number of breast cancers in the presence of screening; T_0, predicted number of breast cancers in the absence of screening. The percentage of deficit (D) breast cancers was calculated as $(T_{0,\ age\ 69/74-100\ years} - T_{1,\ age\ 69/74-100\ years})/T_{0,\ age\ 69/74-100\ years}$. Overdiagnosis was then calculated as the number of excess cancers in the age group 0–69/74 years minus the number of deficit cancers in the age group 69/74–100 years divided by the total number of breast cancers in the absence of screening in women aged 0–100 years.

allow for lead time are crucial to observing a deficit in breast cancer incidence and to calculating overdiagnosis correctly. In several studies, the first years of the screening program were included in the overdiagnosis estimate (21, 23, 29). A relatively large proportion of women will have a prevalence screen in these years, which will increase the number of excess breast cancers. In the worst-case scenario, this could have resulted in overestimation of the overdiagnosis rate by a factor of 4 (Table 3, estimator 1, 1993 vs. 2006). Of course, first (prevalent) screening rounds of women who reach the lower age limit for screening should be included in an overdiagnosis estimate. However, the proportion of women reaching this age will be stable only during steady-state screening.

Several studies based their analyses on the period after implementation of screening but still may not have fully accounted for lead time (5, 9, 22, 27) because they calculated overdiagnosis by using average breast cancer incidence during this phase. However, even during the steady-state phase, overdiagnosis may further drop by a factor of 1.7 (Table 3, estimator 1, 2003 vs. 2006) as the number of women contributing to the deficit in incidence still increases. A compensatory drop in incidence will reach its maximum only if all women in the age group past the screening age had been invited to screening when they were

eligible. Moreover, some tumors may have a lead time longer than 5 years. On the basis of the estimated distribution of the lead time of breast cancer in our study (the best-fitting model assumed a Weibull-distributed lead time with a median of 2 years and a mean lead time of 3.7 years), approximately 20% of all tumors will have a lead time of more than 5 years, and 5% will have a lead time of more than 10 years. Ideally, the lead time of these tumors should be accounted for by calculating overdiagnosis several years after screening has reached the steady-state phase.

Overdiagnosis estimates will be affected by the denominator used to define the population at risk. Various estimators were used in this study, resulting in overdiagnosis estimates differing by a factor of 3.5. By calculating overdiagnosis for women of all ages, we also included women who will never be screened and will not be at risk of overdiagnosis. If overdiagnosis is calculated as a relative risk for women of the screening ages only (5, 21, 22), overdiagnosis could be 1.8 times higher than if women of all ages are included (Table 3, estimator 3 vs. estimator 1, 2006). However, the impact of a screening program is sometimes observed at a later age; by limiting the denominator to the screened age group, the lifetime effect of screening is not given justice. Alternatively, the risk of overdiagnosis can

Table 3. Overdiagnosis Estimates in the Netherlands Using Various Estimators[a]

Phase and Years	Estimator					
	1: $(E - D)/$ $T_{0, \text{ age } 0–100 \text{ years}}$, %	2: $(E - D)/$ $T_{0, \text{ age } 49–100 \text{ years}}$, %	3: $(E - D)/$ $T_{0, \text{ age } 49–69/74 \text{ years}}$, %	4: $(E - D)/$ $T_{1, \text{ age } 49–69/74 \text{ years}}$, %	5: $(E - D)/\text{SD}$, %	6: $T_{1, \text{ 49–69/74 years}}/$ $T_{0, \text{ 49–69/74 years}}$, %
Implementation phase						
1990–1991	1.0	1.4	2.4	2.3	35.4	2.3
1991–1992	6.1	8.2	14.1	12.4	67.4	14.3
1992–1993	9.1	12.2	21.3	17.5	61.5	21.6
1993–1994	11.4	15.2	26.7	21.0	54.7	27.3
1994–1995	10.0	13.3	23.2	18.7	44.5	24.0
1995–1996	9.4	12.4	21.8	17.7	38.2	23.3
1996–1997	8.8	11.6	20.3	16.5	32.6	22.7
1997–1998	5.6	7.3	12.7	11.0	22.1	15.2
Extension phase						
1998–1999	4.9	6.5	9.0	8.3	18.9	9.1
1999–2000	10.0	13.1	18.2	15.4	30.4	18.6
2000–2001	7.4	9.7	13.6	11.8	23.0	14.7
2001–2002	4.7	6.1	8.7	7.8	15.4	10.6
Steady-state phase						
2002–2003	4.7	6.1	8.6	7.7	15.2	11.1
2003–2004	4.9	6.3	8.9	8.0	15.6	11.9
2004–2005	4.3	5.5	7.7	6.9	13.2	11.4
2005–2006	4.4	5.7	7.9	7.0	13.6	11.6
2006–2007	2.8	3.6	5.0	4.6	8.9	9.7

Abbreviation: SD, number of screen-detected cancers.

[a] $E - D$ is the number of excess breast cancers (E) minus the number of deficit breast cancers (D). The excess is calculated as the difference between the modeled number of breast cancers with (T_1) and the predicted number of breast cancers without (T_0) screening in the screened age group; the deficit is calculated as the difference in the predicted number of breast cancers without and the modeled number of cancers with screening in the age groups past the screening age.

also be calculated for women of the screening age and older (7, 19, 20, 23). In this case, the overdiagnosis estimate would be 1.3 times higher than when women of all ages are included (Table 3, estimator 2, 2006). Calculated as the fraction of all breast cancers diagnosed in women of the screening age in a situation with screening, comparable to the estimates by Duffy et al. (6, 25) and Olsen et al. (24), the estimate in the Netherlands would be 4.6% (Table 3, estimator 4, 2006). If overdiagnosis is calculated as a fraction of screen-detected cancers, the estimate would increase by a factor of 3.2 (Table 3, estimator 5, 2006).

A comparable finding was shown by Welch and Black (26), who demonstrated that the overdiagnosis rate in the Malmö trial, previously estimated to be 10% (7), would be 24% if only screen-detected cancers were taken into account. The choice of the denominator will likely depend on the purpose of the overdiagnosis estimate. If the population risks of different screening regimens—for instance, with varying starting ages for screening—are compared, the denominator that includes all diagnosed breast cancers in women aged 0–100 years may be useful. If the main purpose is to inform individual women of their risk of being

overdiagnosed, the denominator that includes women of the screening age and older may be more useful. The fraction of screen-detected cancers overdiagnosed may be relevant in evaluating the performance of a particular screening program or in treatment decisions for DCIS.

Varying overdiagnosis estimates could also be explained by differences in screening characteristics. For instance, the more women are screened, the more likely that an irrelevant tumor is detected. Thus, shorter screening intervals and higher attendance rates may increase the overdiagnosis rate. Overdiagnosis could also be affected by referral or recall practice. If the threshold for diagnostic assessment of small or obscure lesions is higher, fewer of these tumors may be detected or overdiagnosed. The fraction of tumors that are noninvasive may also influence overdiagnosis. In the United States, for instance, 17%–34% of all screen-detected cancers are DCIS (31), whereas, in the Netherlands, this fraction is somewhat lower (16%) (14).

Overdiagnosis estimates will also be affected by the age of the screened group. In the Malmö trial, for instance, the study group was 45–69 years of age at randomization. At 15 years of follow-up, they will be 60–84 years of age. Because

tumors grow slower at older ages, and because mortality from causes other than breast cancer increases with age, such trial-based estimates will be higher than overdiagnosis estimates based on ongoing screening programs that have a constant inflow of women in the lower age limit of screening. Another factor that might bias the overdiagnosis rate is screening of women in the age groups no longer eligible for screening (5, 9) or screening in the control group of a trial (7). For instance, an estimated 24% of women in the control group of the Malmö trial were thought to be screened (19).

Use of mathematical modeling to calculate overdiagnosis has certain limitations. Overdiagnosis estimates will be affected by model assumptions about the natural history of breast cancer. Previous studies showed that model parameters, such as test sensitivity, mean duration of the preclinical phase of cancer, and probability of preclinical DCIS to progress to invasive cancer or to regress, can be interchanged to some extent (32) (R. de Gelder, Erasmus MC, Department of Public Health, unpublished manuscript). This means that, for instance, a model with a higher progression and lower regression rate of preclinical DCIS could simulate observed breast cancer incidence equally well as a model with a lower progression and higher regression rate, provided that test sensitivity, mean duration, and onset rate of breast cancer are adjusted accordingly. This might affect the overdiagnosis rate.

In the present study, the probability of DCIS progression at age 50 years was estimated to be 61% and the probability of regression to be 11%. If we instead assumed that 0% of preclinical DCIS would progress and 96% would regress, the predicted overdiagnosis rate would be 8.1% of all predicted cancers in 2006 (data not shown). This rate is still considerably lower than the overdiagnosis estimates published elsewhere (5, 9, 27). Because the natural history of DCIS is unobservable, no direct evidence exists on the "true" progression and regression rate.

However, indirect evidence suggests that the assumed progression rate in our model is plausible. For instance, follow-up of undertreated DCIS initially misdiagnosed as benign shows that 11%–60% of all DCIS recurs as invasive cancer within 10–20 years (32). Furthermore, basement membrane invasion has been observed in DCIS cases, and microscopy on invasive lesions showed that DCIS was present in 20%–30% of the carcinomas (33). Literature provides no evidence on the fraction of preclinical DCIS that regresses. Because of structural model uncertainties, it would be useful to assess the overdiagnosis rate in one particular screening situation with various collaborating models, such as those in the Cancer Intervention and Surveillance Modeling Network (CISNET) (34). These models share the same input but vary in their structure and assumptions, which reflects the uncertainties about the natural history of breast cancer. Only some of the CISNET models incorporated DCIS or assigned low malignant potential to a fraction of the tumors (34), which would of course affect overdiagnosis estimates.

The present study is limited by the fact that from 2002 onward, the modeled breast cancer incidence is lower than the observed incidence among women of the screening ages (Figure 2). However, increasing the sensitivity of mammography in the model did not result in a substantial increase in the modeled incidence, without affecting the predicted number of interval cancers or clinically diagnosed cancers (14).

The difference between observed and modeled incidence rates should therefore be explained by an increasing trend in background incidence. This could happen when breast cancer incidence increases because of, for instance, a rising prevalence of risk factors such as lower parity, older age at birth of the first child, or obesity. Increasing incidence trends have been observed before implementation of the Dutch screening program and in unscreened women (35, 36). Future modeling efforts should take such background trends into account. Because increases in background incidence are likely to occur in both screened and unscreened women at an approximately similar rate, it is unlikely that our overdiagnosis estimate, which did not include the secular increase in incidence, was affected by poor model fit in recent years. If we conservatively assume that improvement in the sensitivity of mammography would have increased the excess incidence by 10%, the overdiagnosis estimate at steady-state screening would be 3.4%–5.6%. Although the modeled incidence was lower than the actual observed incidence in recent years, we modeled the observed incidence between 1990 and 2002 fairly well. Moreover, the model reproduced the observed incidence decline in women who are no longer screened accurately for all observation years.

Despite several limitations, our approach to calculating overdiagnosis has certain advantages. By using a model, screened and nonscreened populations could be exactly the same, with a similar background risk of developing breast cancer. Studies that, for instance, compare screened and historical comparison groups have the disadvantage that temporal incidence trends may have affected one group but not the other (e.g., by the use of hormone replacement therapy in the late 1990s). Moreover, our approach has the advantage that it is based on observed data from a long-running, population-based mammography screening program with high participation rates (82% in the target population, 95% of previous attendees) that annually targets more than a million women. The observed data on which the model was based include clinically diagnosed breast cancers, screen-detection rates, and interval cancer rates. Natural history parameters, such as lead time, could be estimated from these data. The observations show that the incidence of breast cancer strongly decreases after women have reached the upper age limit for screening. This finding strongly suggests that the risk of overdiagnosis must be smaller than recent estimates of approximately 50% (5, 8, 9, 27).

In conclusion, our estimates of overdiagnosis are substantially lower than those published in recent literature. This discrepancy is most likely related to methodological differences between studies and lack of sufficient follow-up, and partly to differences in screening characteristics and performance. In 2006, the estimated risk of overdiagnosis in the Netherlands ranged between 2.8% of all predicted cancers in women aged 0–100 years in the absence of screening and 9.7% of all predicted cancers in women of the screening age only.

ACKNOWLEDGMENTS

REFERENCES

1. de Koning HJ. Mammographic screening: evidence from randomised controlled trials. *Ann Oncol*. 2003;14(8):1185–1189.
2. International Agency for Research on Cancer. Volume 7: breast cancer screening. In: Vainio H, Bianchini F, eds. *IARC Handbooks of Cancer Prevention*. Lyon, France: IARC Press; 2002.
3. Allgood PC, Warwick J, Warren RM, et al. A case-control study of the impact of the East Anglian breast screening programme on breast cancer mortality. *Br J Cancer*. 2008;98(1):206–209.
4. Waller M, Moss S, Watson J, et al. The effect of mammographic screening and hormone replacement therapy use on breast cancer incidence in England and Wales. *Cancer Epidemiol Biomarkers Prev*. 2007;16(11):2257–2261.
5. Jørgensen KJ, Gøtzsche PC. Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. *BMJ*. 2009;339:b2587. (doi: 10.1136/bmj.b2587).
6. Duffy SW, Agbaje O, Tabar L, et al. Overdiagnosis and overtreatment of breast cancer: estimates of overdiagnosis from two trials of mammographic screening for breast cancer. *Breast Cancer Res*. 2005;7(6):258–265.
7. Zackrisson S, Andersson I, Janzon L, et al. Rate of overdiagnosis of breast cancer 15 years after end of Malmö mammographic screening trial: follow-up study. *BMJ*. 2006;332(7543):689–692.
8. Biesheuvel C, Barratt A, Howard K, et al. Effects of study methods and biases on estimates of invasive breast cancer overdetection with mammography screening: a systematic review. *Lancet Oncol*. 2007;8(12):1129–1138.
9. Morrell S, Barratt A, Irwig L, et al. Estimates of overdiagnosis of invasive breast cancer associated with screening mammography. *Cancer Causes Control*. 2010;21(2):275–282.
10. de Koning HJ, Draisma G, Fracheboud J, et al. Overdiagnosis and overtreatment of breast cancer: microsimulation modelling estimates based on observed screen and clinical data. *Breast Cancer Res*. 2006;8(1):202. (doi:10.1186/bcr1369).
11. de Koning HJ, Boer R, Warmerdam PG, et al. Quantitative interpretation of age-specific mortality reductions from the Swedish breast cancer-screening trials. *J Natl Cancer Inst*. 1995;87(16):1217–1223.
12. Groenewoud JH, Pijnappel RM, van den Akker-Van Marle ME, et al. Cost-effectiveness of stereotactic large-core needle biopsy for nonpalpable breast lesions compared to open-breast biopsy. *Br J Cancer*. 2004;90(2):383–392.
13. Vereniging Integrale Kankercentra VIKC. Incidence data. Utrecht, the Netherlands: VIKC; 2010. (http://www.ikcnet.nl/). (Accessed April 6, 2010).
14. National Evaluation Team for Breast Cancer Screening (NETB). *National Evaluation of Breast Cancer Screening in The Netherlands 1990–2007. Twelfth Evaluation Report*. Rotterdam, Nijmegen, the Netherlands: Erasmus MC, Radboud University Nijmegen Medical Centre; 2009.
15. van den Akker-van Marle ME, Reep-van den Bergh CM, Boer R, et al. Breast cancer screening in Navarra: interpretation of a high detection rate at the first screening round and a low rate at the second round. *Int J Cancer*. 1997;73(4):464–469.
16. Rijnsburger AJ, van Oortmarssen GJ, Boer R, et al. Mammography benefit in the Canadian National Breast Screening Study-2: a model evaluation. *Int J Cancer*. 2004;110(5):756–762.
17. de Gelder R, Bulliard JL, de Wolf C, et al. Cost-effectiveness of opportunistic versus organised mammography screening in Switzerland. *Eur J Cancer*. 2009;45(1):127–138.
18. Beemsterboer PM, de Koning HJ, Looman CW, et al. Mammography requests in general practice during the introduction of nationwide breast cancer screening, 1988–1995. *Eur J Cancer*. 1999;35(3):450–454.
19. Moss S. Overdiagnosis and overtreatment of breast cancer: overdiagnosis in randomised controlled trials of breast cancer screening. *Breast Cancer Res*. 2005;7(5):230–234.
20. Puliti D, Paci E. The other side of technology: risk of overdiagnosis of breast cancer with mammography screening. *Future Oncol*. 2009;5(4):481–491.
21. 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(6):R68. (doi:10.1186/bcr1625).
22. Jørgensen KJ, Zahl PH, Gøtzsche PC. Overdiagnosis in organised mammography screening in Denmark. A comparative study. *BMC Womens Health*. 2009;9:36. (doi:10.1186/1472-6874-9-36).
23. Paci E, Warwick J, Falini P, et al. Overdiagnosis in screening: is the increase in breast cancer incidence rates a cause for concern? *J Med Screen*. 2004;11(1):23–27.
24. Olsen AH, Agbaje OF, Myles JP, et al. Overdiagnosis, sojourn time, and sensitivity in the Copenhagen mammography screening program. *Breast J*. 2006;12(4):338–342.
25. 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(1):25–30.
26. Welch HG, Black WC. Overdiagnosis in cancer. *J Natl Cancer Inst*. 2010;102(9):605–613.
27. 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(7445):921–924.
28. 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(5):842–847.
29. Martinez-Alonso M, Vilaprinyo E, Marcos-Gragera R, et al. Breast cancer incidence and overdiagnosis in Catalonia (Spain). *Breast Cancer Res*. 2010;12(4):R58. (doi:10.1186/bcr2620).
30. Duffy SW, Lynge E, Jonsson H, et al. Complexities in the estimation of overdiagnosis in breast cancer screening. *Br J Cancer*. 2008;99(7):1176–1178.
31. Kerlikowske K. Epidemiology of ductal carcinoma in situ. *J Natl Cancer Inst Monogr*. 2010;2010(41):139–141.
32. McCann J, Treasure P, Duffy S. Modelling the impact of detecting and treating ductal carcinoma in situ in a breast screening programme. *J Med Screen*. 2004;11(3):117–125.
33. Feig SA. Ductal carcinoma in situ. Implications for screening mammography. *Radiol Clin North Am*. 2000;38(4):653–668, vii.

34. Cronin KA, Feuer EJ, Clarke LD, et al. Impact of adjuvant therapy and mammography on U.S. mortality from 1975 to 2000: comparison of mortality results from the CISNET breast cancer base case analysis. *J Natl Cancer Inst Monogr.* 2006; 36:112–121.

35. Louwman WJ, Voogd AC, van Dijck JA, et al. On the rising trends of incidence and prognosis for breast cancer patients diagnosed 1975–2004: a long-term population-based study in southeastern Netherlands. *Cancer Causes Control.* 2008; 19(1):97–106.

36. Otten JD, Broeders MJ, Fracheboud J, et al. Impressive time-related influence of the Dutch screening programme on breast cancer incidence and mortality, 1975–2006. *Int J Cancer.* 2008;123(8):1929–1934.